THE UNIVERSITY
of EDINBURGH

The Contribution of *Lsh* to DNA Methylation During Neurodevelopment

Leanne Beth Duthie

THE UNIVERSITY *of* EDINBURGH

Presented for the Degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis was written by me and all the work it presents was performed by myself unless otherwise stated.

This work has not been submitted for any other degree or professional qualification.

Leanne Duthie

Acknowledgements

This project would not have been possible without the help and support of so many people.

I would like to first thank Stephen Lawrie, and all involved in setting up the PsySTAR scheme, in putting your trust in me and funding this project.

A big thank you to Richard Meehan for dreaming up this project and welcoming me into the lab despite my lack of experience. Your enthusiasm and calming words of advice have kept me motivated throughout this project. Also thank you to my second supervisor Mandy for helping me find a research topic in the first place and for introducing me to people who have been very helpful at different stages during this project.

The biggest thank you should go to my immediate supervisor Donncha for teaching me pretty much everything! I really appreciate all the work you put into the bioinformatic analysis and for patiently trying to explain it to a medic. I would also like to thank you for all your support when I announced I was pregnant and for all the parenting tips.

All members of the Meehan lab have supported me in various ways but a special mention should go to Heidi for all her cell culture help and for always answering my stupid lab questions with great patience. Your enthusiasm is infectious and my time in the lab would not have been so enjoyable without you there. Another special mention goes to Ailsa who taught me most of the lab techniques and put up with me shadowing her every move. This project would not have been possible without all the work and expertise of Ian Adams and David Read in generating the mouse model.

I would also like to thank my parents for all the support they have given me throughout my (prolonged) education and for all the babysitting! Thank you to my brother Craig for his superpowers in grammar and Claire for your helpful comments. Thank you also to Geo for all your support, advice and encouragement throughout my career. Murron, your games of peek-a-boo have helped me remember what life is really all about.

Finally a very special thank you to my husband Chris. I know that none of this would have been possible without you.

Abstract

DNA methylation is considered one of the oldest epigenetic modifications and is associated with repression of transcriptional activity. Embryonic development sees dynamic changes in this mark. Initially, widespread demethylation occurs with the establishment of pluripotency in the developing zygote. This is then followed by a wave of *de novo* methylation in the implanting blastocyst, coincident with lineage specification, which continues until tissue specific patterns of DNA methylation are laid down.

Disruptions in DNA methylation, and mutations in genes that form part of the DNA methylation machinery, are associated with neurodevelopmental disorders. Therefore, understanding the mechanisms of DNA methylation during neurodevelopment is key if we are to understand the pathology underlying these conditions. Investigating this has, thus far, proven difficult due to the lack of available models that survive into adulthood. Therefore the consequences of dysregulated DNA methylation during neurodevelopment on the mature brain are unknown.

The DNA methyltransferase enzymes are responsible for the deposition of DNA methyl marks, although less is known about the cofactors required for their targeting and activity. *Lsh* (lymphoid specific helicase), a chromatin remodeller, has been described to play a key role in *de novo* methylation during development. The importance of this protein is highlighted by the fact that knockout mice die within a few hours of birth. Furthermore, whilst I was pursuing this project, *Thijsen et al* described mutations in this gene being causative for Immunodeficiency, Centromeric instability and Facial anomalies syndrome (ICF), in which a large proportion of patients suffer from intellectual disability (Thijssen et al., 2015).

At the beginning of this project, nothing was known about the role of *Lsh* during neurodevelopment, nor why mutations in this gene should result in neurological defects. This thesis aims to determine the contribution of *Lsh* to DNA methylation during neurodevelopment and to investigate the consequences of its absence on the mature brain.

In order to investigate the roles of *Lsh* at early stages of neurogenesis I made use of an *in vitro* neurodifferentiation system to investigate the differentiation of *Lsh*$^{-/-}$ mouse embryonic stem cells down neural lineage. This revealed an enhanced differentiation of these cells down neural lineage compared to wild type. Genome-wide methylome analysis uncovered a key role for *Lsh* in establishing appropriate DNA methylation at repetitive sequences during this developmental window.

Previous papers, investigating selected loci, have suggested a role for *Lsh* in methylating promoters of single copy genes thereby controlling their transcription. By investigating transcription genome–wide, I noted a key role for *Lsh* in regulating genes associated with

developmental processes. This regulation was not, however, related to their promoter methylation status. This led me to investigate the regulation of the Polycomb system in these cells, another key repressive epigenetic system involved in developmental gene transcriptional regulation.

This analysis uncovered wide-scale redistribution of the polycomb mark H3K27me3 from target genes to hypomethylated repeats, similar to what is seen in other hypomethylated models. Whilst this could account for the dysregulation of developmental gene expression, I also discovered preliminary evidence of a role for *Lsh* in regulating transcription at these sites independent of its role in DNA methylation.

To investigate the role of *Lsh in vivo,* I used a novel *Nestin-Cre* knockout mouse model. This new targeted mouse model, unlike previous models, survived to adulthood. This allowed investigation of the effects that disruption of DNA methylation during development has on the mature brain. This mouse displayed no gross changes in brain morphology or behaviour. DNA methylation analysis revealed gross hypomethylation at repeat sequences particularly in the knockout cerebellum. A consequence of this was aberrant repeat transcription. RNA-Seq revealed activation of innate immune response genes. This led me to propose a model by which aberrant repeat transcription results in an immune response due to cellular detection of cytoplasmic double stranded DNA generated via reverse transcription. This opens up exciting new avenues of research into the underlying pathology of the neurological deficits seen in ICF syndrome and potential therapeutic options.

Lay Summary

Each cell in the body contains the same DNA, yet a brain cell is very different from a liver cell for example. This is due to the fact that different genes will be switched on or off in different cell types. One of the ways this is achieved is through adding different chemical tags to the DNA. One important example is the addition of a methyl group onto the DNA. This DNA methylation plays an important role in switching off genes and so called “junk DNA”, that is sections of DNA that do not contain any genes.

After fertilisation of an egg, the DNA methylation marks from both parents are erased to produce a “clean slate”. During development of the embryo there is then a wave of DNA methylation ensuring the correct genes are switched off in each cell type. As a psychiatrist I am particularly interested in this process, as defects in it are associated with developmental disorders of the brain. Yet we still do not fully understand what factors are important in DNA methylation or how problems in it can result in brain disorders.

One protein known to be of importance in DNA methylation during development is *Lsh*, however we do not know the role of this protein specifically in brain development. Mutations in *Lsh* can cause ICF syndrome (Immunodeficiency, Centromeric instability and Facial anomalies) in which the majority of patients suffer from intellectual disability.

During this project I aimed to determine the role of *Lsh* in DNA methylation during brain development. To do this I used mouse embryonic stem cells lacking *Lsh,* which models the early stage of embryonic development just before the wave of DNA methylation occurs. I then stimulated them to develop into early brain cells. I also used a mouse model lacking *Lsh* in brain cells only. I found that *Lsh* was required for normal DNA methylation during the early stages of brain development, particularly at DNA sequences called repetitive elements which make up the majority of so called “junk DNA”. This resulted in a failure to switch off these elements. This is particularly important as activation of repetitive sequences can have many detrimental consequences.

Another exciting discovery was the activation of the immune system in the brains of mice lacking *Lsh*, which I propose is a result of the activation of repetitive elements. This could be a potential explanation for why patients with ICF suffer from intellectual disability and certainly opens up further avenues of research in this area.

To conclude, the work presented in this thesis advances our understanding of DNA methylation mechanisms during brain development. Furthermore, I have uncovered potential pathways which may explain the occurrence of intellectual disability in patients with ICF syndrome warranting further investigation.

Contents

List of Figures

List of Tables / Lists

Abbreviations

5caC -	5'-carboxylcytosine
5fc -	5'-formylcytosine
5hmC -	5'-hydroxymethylcytosine
5mC -	5'-methylcytosine
ADCA-DN -	Autosomal Dominant Cerebellar Ataxia, Deafness and Narcolepsy
AEU -	Arbitrary expression units
AGS -	Aicardi-Goutieres Syndrome
ATP -	Adenosine triphosphate
BMP4 -	Bone morphogenic protein 4
bp -	Base pairs of DNA
BSA -	Bovine serum albumin
Cdca7 -	Cell division cycle associated 7
CDK -	Cyclin dependent kinase
cDNA -	Copy DNA
CDR -	Cellular demethylation and remethylation
CGI -	CpG island
ChIP -	Chromatin immunoprecipitation
CHIRCC -	CDCA7-HELLS ICF-related nucleosome remodelling complex
CpG -	Cytosine base directly preceding a guanosine base
CTCF -	CCCTC-binding factor
DAPI -	4',6-diamidino-2-phenylindole
DDM1 -	Decrease in DNA Methylation 1
dH20 -	Distilled water
DMEM -	Dulbecco's modified eagle media
DNA -	Deoxyribonucleic acid
DNMT -	DNA methyltransferase
DTT -	Dithiothreitol
E -	Embryonic day
EED -	Embryonic ectoderm development
ERRBS -	Enhanced reduced representation bisulphite sequencing
ERV -	Endogenous retrovirus
ESC -	Embryonic stem cell
EtBr -	Ethidium bromide
EZH -	Enhancer of zeste
FACS -	Fluorescent activated cell sorting

FCS - Fetal calf serum
FITC - Fluorescein isothiocyanate
FRT - Flippase recognition target site
FSC-A - Forward scatter area
FSC-H - Forward scatter height
GO - Gene ontology
H3K27 - Histone H3 lysine 27
H3K4 - Histone H3 lysine 4
H3K9 - Histone H3 lysine 9
HAT - Histone acetyltransferase
HCPs - High density CpG promoter
HDAC - Histone deacetylase
Hells - Helicase, lymphoid specific
HGU - Human Genetics Unit
HMT - Histone methyltransferase
Hox - Homeobox
HP1 - Heterochromatin protein 1
HSAN - Hereditary and Sensory Autonomic Neuropathy
IAP - Intercisternal A particle
ICF - Immunodeficiency, centromere instability and facial anomalies syndrome
ICM - Inner cell mass
ICP - Intermediate density CpG promoter
ICR - Imprinting control region
IP - Immunoprecipitation
iPSC - Induced pluripotent stem cell
kb - Kilobase pairs of DNA
KO - Knockout
L1 - Long interspersed nuclear element
LAD - Lamin B1 attachment domain
LB - Luria-Bertani
LC-MS - Liquid chromatography mass spectrometry
LCP - Low density CpG promoter
LIF - Leukaemia Inhibitory factor
LINE - Long interspersed nuclear element
lncRNa - Long non-coding RNA
Lsh - Lymphoid specific helicase
LTR - Long terminal repeat
MACS - Magnetic activated cell sorting

MAP -	Methyl-CpG-binding affinity purification
MBD -	Methyl-CpG binding domain
MECP -	Methyl-CpG binding protein
MeDIP -	Methyl-DNA immunoprecipitation
MEF -	Mouse embryonic fibroblast
mESC -	Mouse embryonic stem cell
ncRNA -	Non-coding RNA
NEP -	Neuroepithelial
NOMe -	Nucleosome occupancy and methylome
NPC -	Neural progenitor cell
ORF -	Open reading frame
PASG -	Proliferation associated SNF2-like gene
PcG -	Polycomb group protein
PCR -	Polymerase chain reaction
PFA -	Paraformaldehyde
PGC-	Primordial germ cell
pol -	RNA polymerase
PRC -	Polycomb repressor complex
PSA-NCAM -	Polysialylated-neural cell adhesion molecule
qRT-PCR -	Quantitative reverse transcriptase polymerase chain reaction
RGC -	Radial glial cell
RING -	Really interesting new gene
RNA -	Ribo-nucleic acid
RPM -	Rotations per minute/reads per million
RRBS -	Reduced representation bisulphite sequencing
SASP -	Senescence-associated secretory phenotype
SDS -	Sodium dodecylsulphate
SEM -	Standard error of the mean
Seq -	Sequencing
SET -	Su(var)3-9, Enhancer of zeste and Trithorax protein domain
SINE -	Short interspersed nuclear element
siRNA -	Short interfering RNA
SMARCA -	SWI/SNF related matrix associated actin- dependent regulator of chromatin Subfamily a member 6
SRA -	SET and RING associated domain
Suz -	Suppressor of zeste
SWI/SNF -	Switch/sucrose-non-fermenting
TAE -	Tris-acetate-EDTA

TBE -	Tris-borate-EDTA
TDG -	Thymine DNA glycosylase
TE -	Tris EDTA
TET -	Ten-eleven translocation
TF -	Transcription factor
TSS -	Transcription start site
Tuj1 -	Neuron-specific class 111 β-tubulin
UCSC -	University of Santa Cruz
WGBS -	Whole genome bisulphite sequencing
WT -	Wild type
Xa -	Active X chromosome
Xi -	Inactive X chromosome
Xist -	X-inactive specific transcript

Chapter 1. Introduction

1.1 Epigenetics

Every cell in a mammalian organism contains an identical genome, with a few exceptions in the immune and reproductive systems. The underlying genetic sequence therefore cannot account for differences in cellular phenotype. Similarly, during development hundreds of cell types arise from a single pluripotent cell and these progeny will display a different phenotype dependent upon their stage of development. It is the differences, therefore, in gene expression that account for these differences in cellular phenotype. Epigenetics is the study of certain mechanisms important in this transcriptional control.

Epigenetics was a term coined by the embryologist Conrad Waddington in the 1940's who defined it as "the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being" (Waddington, 1968, 2012). Waddington's concept of the epigenetic landscape was concerned with the decisions made by a cell during differentiation. He likened a pluripotent cell to a marble, rolling down a hilly landscape in which it takes specific paths leading to different cell fates. Each time it takes a decision to enter a certain path on rolling down the hill its final destination options become more restricted; beautifully describing the stages of cell differentiation.

This original description referred to all the molecular pathways influencing expression of a genotype. However, over the years this term has been co-opted by geneticists and molecular biologists, resulting in more narrowed definitions to describe mechanisms that influence gene transcription. This led to the generally accepted modern definition of epigenetics as "heritable changes in gene expression which are not due to changes in the underlying DNA sequence"(Bird, 2007). However, the idea of heritability does not hold true in all situations, for example there are alterations in epigenetic marks and gene transcripts in response to activity in post-mitotic neurons (Guo et al., 2011; Lubin et al., 2008; Ma et al., 2009; Yu et al., 2015). There is also a growing recognition that epigenetic states may be a consequence of transcriptional activity rather than a cause (Henikoff and Shilatifard, 2011; Sproul and Meehan, 2013). Therefore, perhaps a more relevant definition is "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states"(Bird, 2007).

Today, epigenetics most frequently, but not exclusively, refers to covalent modifications on DNA and histones that may alter their biological properties. This network of chemical tags can be referred to as the "epigenome". The role of non-coding RNA's in regulating gene activity are also now part of the epigenetics canon (Deichmann, 2016).

The role of epigenetics in somatic cells is very interesting, particularly in the field of neuroepigenetics where it may offer up exciting insights into how the brain functions in health and disease. However, it is during embryogenesis that we see the most dramatic alterations in the epigenome, which occur in parallel with the orchestration of cell differentiation during the development of a highly complex organism.

During this research project, I have attempted to delve deeper into our understanding of epigenetic mechanisms during the early stages of neurodevelopment, with a focus on the most well studied epigenetic modification - that of DNA methylation. Specifically, I have studied the contribution of the chromatin remodeller, lymphoid specific helicase (*Lsh*), to DNA methylation reprogramming during neurodevelopment using both a cell culture and a mouse model system.

In this chapter, I will discuss DNA methylation and its dynamics during development. I will also review the known or speculated roles of *Lsh* into its establishment. The importance of research into this area is highlighted with a discussion of neurodevelopmental disorders associated with perturbation of DNA methylation pathways during development.

1.2 Chromatin Organisation

In order to understand the effects of different epigenetic modifications, one has to have an understanding of how DNA is organised within a cell.

In mammalian cells an astonishing 2 meters of DNA has to be packaged into a nucleus only 6μm in diameter. This is achieved by compacting the DNA into chromatin. The basic unit of chromatin is the nucleosome, which is comprised of 146bp of DNA wrapped around a globular histone octamer composed of the 4 core histone proteins H2A, H2B, H3 AND H4 (Burlingame et al., 1985; Luger et al., 1997).

The positioning of these nucleosomes will determine the degree of chromatin compaction. Densely compacted chromatin is referred to as heterochromatin, exemplified by constitutive heterochromatin at mouse centromeric regions, which results in DNA being inaccessible and hence may prevent binding of transcription factors (TFs). Genes are therefore transcriptionally silenced at heterochromatin. Alternatively, compaction can be relaxed to form euchromatin in which the DNA is accessible and amenable to transcription factor binding, and therefore can be actively transcribed. This lower order chromatin structure has to be dynamic: particularly during development where extensive remodelling occurs to allow vast changes in gene expression. A summary of chromatin organisation and the key epigenetic marks involved is shown in Figure 1.1.

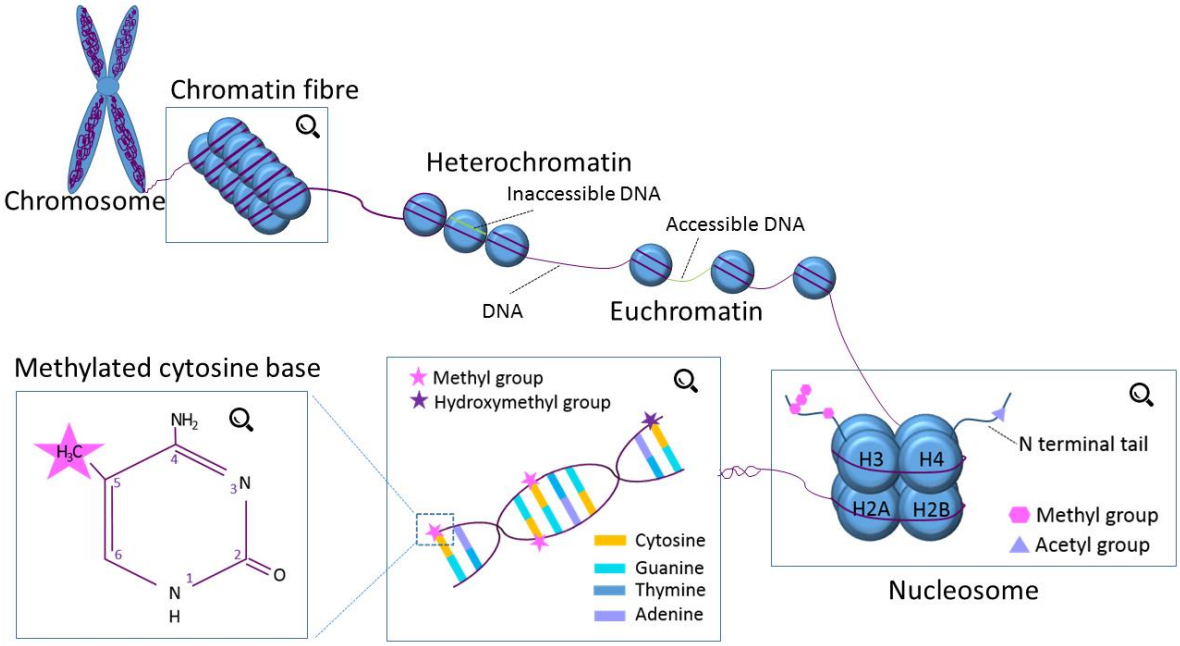


Figure 1.1. Chromatin Organisation. Diagrammatic representation of how DNA is compacted into chromatin. Chromosomes are composed of tightly compacted chromatin fibres. The basic unit of chromatin is the nucleosome. This consists of 146bp of DNA wrapped around a globular histone octamer composed of the 4 core histone proteins H2A, H2B, H3 AND H4. These histone proteins have N terminal tails which are subject to modifications including methylation and acetylation. When nucleosomes are tightly compacted DNA is inaccessible and this is referred to as heterochromatin. In contrast transcriptionally permissive euchromatin is characterised by less nucleosome dense regions. DNA methylation is an important epigenetic mark involved in maintaining a heterochromatic state. The methyl group is deposited on the 5th carbon position of the pyrimidine ring of cytosine bases adjacent to guanine by the DNA methyltransferases. Cytosine bases can also be hydroxymethylated which may be part of a demethylation process.

There are also higher order levels of chromatin organisation within the nucleus. For example, heterochromatin tends to be located towards the nuclear periphery and euchromatin clustered centrally (Rae and Franke, 1972). Furthermore, regulatory gene regions can be brought into contact with target genes hundreds of kilobases away, via the formation of chromatin loops, to regulate transcription (Carter et al., 2002; Tolhuis et al., 2002). This leads to the view of a network of chromatin based mechanisms, in which epigenetics play a constituent part, which regulate gene expression. Development itself is underpinned by essential signalling and transcription factor (TF) networks that are reinforced by epigenetic pathways that ultimately define cell identity in differentiated cells (Bogdanovic and Lister, 2017; Peter and Davidson, 2016).

1.2.1 Histone Modifications

Post-translational modifications on histone proteins form part of the epigenome. They have the ability to alter chromatin structure by directly affecting nucleosome-DNA electrostatic interactions, or indirectly by binding or preventing binding of chromatin modifying proteins. These modifications occur predominantly on the histone N terminal tails which extend out from the nucleosome, although they do also occur on the globular core domains (Figure 1.1). N terminal tails can be subject to many modifications including: methylation, acetylation, ubiquitination, citrullination, phosphorylation, biotinylation and sumoylation (Bannister and Kouzarides, 2011). The most extensively studied modifications are those of methylation and acetylation at lysine residues.

Lysine residues can be mono-, di- or tri-methylated (me1, me2, me3) (Lorenzo and Bedford, 2011; Martin and Zhang, 2005). These modifications are deposited and removed by groups of enzymes known as histone methyltransferases (HMTs) and histone demethylases (HDMTs). These enzymes target specific lysines and methylate them to a specific extent (Collins et al., 2005). The transcriptional effect of histone methylation depends upon which residue is methylated and to what extent. Similarly, histone demethylases show high specificity to their substrate and its degree of methylation (Tsai et al., 2014).

Numerous different protein domains have been found to recognise lysine methylation including PHD fingers, chromodomains, PWWP, MBT and Tudor domains (Champagne and Kutateladze, 2009; Kim et al., 2006) allowing the binding of many different proteins and complexes which then act downstream to affect chromatin structure. As well as recruiting chromatin associated factors, lysine methylation can also maintain chromatin structure by preventing binding of complexes which would lead to remodelling (Zegerman et al., 2002).

Histone acetylation at lysine residues was linked to transcriptional activity many years ago (Allfrey et al., 1964). Histone acetyl transferases (HATs) and histone deacetylases (HDACs) catalyse its deposition and removal (Krajewski and Becker, 1998). These enzymes show much less specificity than their HMT/HDMT counterparts and are often found in large complexes where they are present with other proteins capable of altering transcription (Yang and Seto, 2007). Unlike lysine methylation, acetylation can directly affect chromatin structure as the deposition of an acetyl group neutralises the lysine's positive charge; weakening affinity between histones and the DNA backbone, which is negatively charged, thereby decompacting chromatin (McGhee and Felsenfeld, 1980). As well as this direct effect, acetylated lysines are also targeted by proteins containing bromodomains which can recruit other remodelling complexes (Zeng et al., 2010).

In reality, however, one histone mark does not equate to one transcriptional outcome. Rather this depends upon the pattern of marks and how they interact with each other. This "histone

code", or signature, is a complex system given the vast number of patterns that can be conceived (Yuan et al., 2011).

1.2.2 Formation of euchromatin

As mentioned above, histone acetylation is a key regulator of euchromatin formation. Hence, this modification is enriched at enhancers and gene promoters (Wang et al., 2008). Certain histone methylation marks are also associated with euchromatin. For example methylation of H3K4 is associated with transcriptionally active regions with H3K4me1 and H3K4me2 being found at enhancers and H3K4me2 and H3K4me3 at the promoters of active genes (Hon et al., 2009; Schneider et al., 2004).

Amongst the complexes recruited to the sites of these histone modifications are chromatin remodellers. These are large multi-protein complexes, capable of remodelling nucleosomes to create a relaxed euchromatic environment permissive for transcription. Remodelling enzymes are subdivided into 4 main families: SWI/SNF (switch/sucrose-non-fermenting), ISWI (imitation switch), CHD (chromodomain-helicase-DNA binding) and INO80 (inositol requiring 80)(Tyagi et al., 2016). All contain a highly conserved ATPase domain and it is with the energy derived from ATP hydrolysis that these enzymes have the power to restructure chromatin by mobilising, ejecting, exchanging or assembling nucleosomes. The other variable domains play a role in recruitment and regulation of the enzyme providing it with its unique biological function.

There are various ways these enzymes can remodel nucleosomes to alter transcription. One way in which the nucleosome can be modified is through the incorporation of histone variants. With the exception of H4, all other histones have been shown to have non-canonical variants whose presence can alter nucleosome positioning and local chromatin structure (Thakar et al., 2009). For example, H2A can be exchanged for H2A.Z by the nucleosome remodeller SWR1, enhancing transcription at promoters and enhancers (Barski et al., 2007; Schones et al., 2008). Other methods include mobilisation of the nucleosomes along the DNA, or complete eviction to clear them from actively transcribed regions allowing access of RNA Polymerase II (Meersseman et al., 1992; Whitehouse et al., 1999) (Figure 1.2).

Although the focus here has been on the role of chromatin remodellers in transcriptional activation, they can also act to create more tightly compacted nucleosomes with some capable of nucleosome assembly and hence can be associated with transcriptional repression (Murawska and Brehm, 2011).

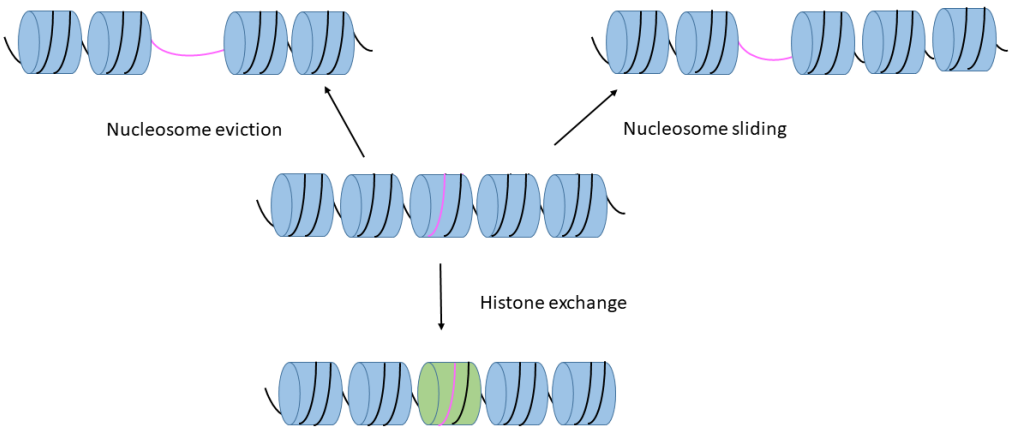


Figure 1.2. Nucleosome remodelling. ATP-dependent chromatin remodellers utilise the energy derived from ATP hydrolysis to alter access to DNA by evicting nucleosomes, causing sliding of nucleosomes along the DNA or by exchanging histone octamers. The diagram illustrates nucleosomes composed of 146bp of DNA wrapped around a globular histone octamer.

1.2.3 Formation of Heterochromatin

Heterochromatin can be further subdivided into facultative or constitutive. The term facultative is derived from “facultus”, the Latin for opportunity. It therefore describes areas of heterochromatin that have the opportunity to become euchromatic, and hence transcriptionally active in response to developmental or environmental cues (Grewal and Jia, 2007). The most prominent histone modification at facultative heterochromatin is H3K27me3, which is deposited by the Polycomb group proteins (PcG) (Cao et al., 2002).

Of particular interest during mammalian development is the presence of both the repressive H3K27me3 mark and the active H3K4me3 mark at a set of genes with a high CG content; so called CpG islands (CGIs) (Bernstein et al., 2006). These regions are referred to as “bivalent” chromatin enriched at developmentally important genes. It is thought that this chromatin state allows the gene to remain poised for either activation or permanent silencing upon differentiation (Mikkelsen et al., 2007).

Constitutive heterochromatin is highly condensed and is required to protect certain sequences, such as repetitive elements, from the transcriptional machinery maintaining their repressed state. The bulk of constitutive heterochromatin forms at pericentromeric regions, which are gene poor and rich in repeat sequences. A hallmark of constitutive heterochromatin is H3K9me3, which binds the protein HP1 via its chromodomain. This is proposed to create a scaffold for recruitment of HMTs and HDACs to create a positive feedback loop to maintain a heterochromatic state (Lachner et al., 2001). This positive feedback loop can also promote the spreading of heterochromatin to neighbouring regions, forming large areas of heterochromatic domains that may have consequences for the

transcription of nearby genes. The ability of heterochromatin to promote its own spreading necessitates the presence of mechanisms to restrict its expansion, in order to prevent erroneous gene silencing. Such mechanisms include DNA sequences that are unfavourable for nucleosome assembly and recruitment of proteins with anti-silencing activities (Allshire and Madhani, 2018).

Somewhat counterintuitively, transcripts from heterochromatic regions may also play a role in heterochromatin formation. In yeast, RNA from pericentromeric transcripts is cleaved to form short-interfering RNAs (siRNAs), which can recruit transcriptional silencing complexes and HMTs (Lippman and Martienssen, 2004).

Another key epigenetic modification associated with the formation of heterochromatin is DNA methylation.

1.3 DNA Methylation

Considered the oldest in evolutionary terms and most stable of epigenetic modifications is DNA methylation. It is believed that this mark evolved as a defence mechanism against foreign DNA integration into the host genome but was then adopted during evolution for transcriptional regulation.

In mammals DNA methylation is mainly found on the 5th carbon position of the pyrimidine ring of cytosine bases adjacent to guanine in a 5'-CpG-3' dinucleotide context, with around 70% of these positions being methylated (Cooper and Krawczak, 1989). Its presence in the genomic landscape exerts a key repressive action on transcriptional regulation (Kass et al., 1993; Keshet et al., 1985). There are, of course, exceptions to this dogma, as advanced sequencing techniques have allowed the identification of non-CpG methylation, and it has become clear that the effects of DNA methylation can vary depending on context (Lister et al., 2009). There are also controversial reports of methylation at adenine bases in mammals (Jeltsch et al., 2018; Lentini et al., 2018; Wu et al., 2016; Xiao et al., 2018).

1.3.1 Key roles of DNA methylation

The key and most established role for DNA methylation is that of maintaining stable transcriptional repression as highlighted in the three biological contexts below: genomic imprinting, X chromosome inactivation and suppression of repetitive elements.

1.3.1.1 Genomic Imprinting

Imprinted genes are those that are expressed from one allele only. Which gene copy is silenced, and which is expressed, depends upon their parent of origin, and is determined by differential DNA methylation established during gametogenesis, which is maintained in the resultant offspring (Ferguson-Smith, 2011). Several human diseases can occur when imprinting is disturbed.

In some instances, it is DNA methylation at a gene promoter that directly determines its silencing, and hence monoallelic expression. Often, however, transcriptional control is more complex and DNA methylation at imprinting control regions (ICRs) can influence expression of distal genes or gene clusters by different mechanisms.

One method by which DNA methylation achieves monoallelic expression is through insulators, which are elements that can block enhancer and promoter interactions. An example of this is the paternally methylated ICR 2kb upstream of *H19* and 80kb downstream of *Igf2*. This ICR on the maternal allele is free to bind the insulator protein CTCF, which blocks the access of *Igf2* to downstream enhancers shared with *H19,* allowing exclusive unrivalled access of *H19* to these enhancers. In contrast, methylation at this site on the paternal allele prevents CTCF binding allowing access of *Igf2* to the enhancers. Thereby facilitating monoallelic expression of *Igf2* from the paternal chromosome and *H19* from the maternal (Hark et al., 2000; Thorvaldsen et al., 1998).

Another method is by influencing expression of long non-coding RNAs (lncRNas). For example in mouse, a maternally methylated ICR at the promoter of a lncRNA transcript, *Air*, causes repression on the maternal chromosome. In contrast on the paternal chromosome lacking this methyl mark *Air* is expressed and acts to repress a neighbouring imprinted gene cluster causing these genes to be expressed from the maternal chromosome in a monoallelic manner (Sleutels et al., 2002). The method by which *Air* does this appears to be via the recruitment of repressive histone modifying proteins (Nagano et al., 2008).

1.3.1.2 X Chromosome Inactivation

Mammalian female cells contain 2 copies of the X chromosome. In order to prevent a doubling in expression of X linked genes in females, one of these copies is randomly inactivated so that in female somatic cells there is a single active X (Xa) and inactivated X (Xi). This inactivation occurs in the epiblast and is maintained through all further cell divisions (Migeon, 2016).

Initial events in X chromosome inactivation actually do not involve DNA methylation. This is evidenced by the fact that embryos lacking the de novo methyltransferases retain the ability to establish it (Sado et al., 2004). The initial silencing instead occurs via the action of the lncRNA *Xist*, expressed from the Xi, which then coats this chromosome recruiting other

silencing complexes, resulting in the deposition of repressive histone marks and the formation of heterochromatin (Chen et al., 2016; Chu et al., 2015; Clemson et al., 1996).

DNA methylation then becomes important in maintaining this silenced state as deletion of the maintenance DNA methyltransferase, *Dnmt1*, in mouse somatic cells results in decompaction and reactivation of the Xi genes (Sado et al., 2000). Increased DNA methylation can be seen over silenced gene promoters on the Xi (Sharp et al., 2011). Furthermore methylation of the *Xist* promoter itself on the Xa prevents aberrant transcription from this chromosome (Sado et al., 2004).

1.3.1.3 Suppression of Repetitive Elements

The final key role for DNA methylation is in the suppression of repetitive elements (Crichton et al., 2014). Recent estimates suggest up to 70% of the human genome is repetitive or derived from repeats (de Koning et al., 2011). Repression of these elements is necessary to maintain genome stability.

Repeat sequences are classed as tandem or interspersed. Satellite sequences are tandemly repeated non-coding clusters, which overlap the centromere and pericentromeric regions and play an essential role in centromere function (Garrido-Ramos, 2017). DNA methylation of satellite elements plays a key role in formation of heterochromatin at centromeres. Hypomethylation of these regions, as seen in human diseases such as Immunodeficiency, centromeric region instability, facial anomalies (ICF) syndrome and cancer or induced by experimental means, results in chromosomal instability with mis-segregation during mitosis (Costa et al., 2006; Hassan et al., 2001; Prada et al., 2012; Widschwendter et al., 2004).

Transposable elements (TEs) are the remnants of exogenous retroviral infection of the ancestral germline. They are interspersed throughout the genome and can be classed into those that require reverse transcription in order to transpose and those that do not. The latter are known as DNA transposons composing about 1% of the mammalian genome (Waterston et al., 2002). These elements encode a transposase which is required for their excision and insertion into the genome via a “cut and paste” mechanism. These elements are no longer capable of mobilising having not been active for the last 37 million years of primate evolution (Pace and Feschotte, 2007).

The other class of TEs are termed retrotransposons, composed of long terminal repeats (LTRs) and non-LTRs, which make up 40% of the mammalian genome (Waterston et al., 2002). These elements do not encode a transposase, but instead rely on the process of reverse transcription for their mobility.

LTRs are also known as endogenous retroviruses (ERVs). They are flanked at their 5’ and 3’ ends by long terminal repeats, which drive RNA transcription and maturation from the endogenous retrovirus. These ERVs also contain *Gag* and *Pol* genes, which encode proteins

involved in retrotransposition. Due to high rates of homologous recombination between flanking LTRs, the intervening ERV sequence is often deleted, resulting in the presence of many solo LTRs in the genome (Belshaw et al., 2007).

ERV retrotransposition occurs within the mouse genome via a “copy and paste” mechanism, whereby the transcribed RNAs are transported to the cytoplasm and packaged into virus-like particles by *Gag* and reverse transcribed by *Pol*, the dsDNA is then transported back into the nucleus where the integrase activity of *Pol* can insert it back into the genome (Mager and Stoye, 2015). In humans there are no known ERVs still capable of retrotransposition.

Non-LTRs are subdivided into long and short interspersed nuclear elements (LINEs and SINEs). The majority of LINEs are of subclass LINE1 (L1s) which are the only active repeat capable of autonomous retrotransposition in humans. The 5’ UTR contains a promoter with both sense and antisense activity and the 3’UTR contains a poly(A) tail. LINE1s also contain 2 important open reading frames (ORF1 and ORF2). ORF1 encodes an RNA binding protein with nucleic acid chaperone activity, and ORF2 encodes a protein with reverse transcriptase and endonuclease activity allowing integration of newly generated cDNA into a new site in the genome (Macia et al., 2015; Mager and Stoye, 2015). Most L1s are 5’ truncated probably due to incomplete reverse transcription during retrotransposition (Waterston et al., 2002). SINEs have to hijack LINE1 machinery in order to retrotranspose. A summary of repeat classes found in the mouse genome is shown in Figure 1.3.

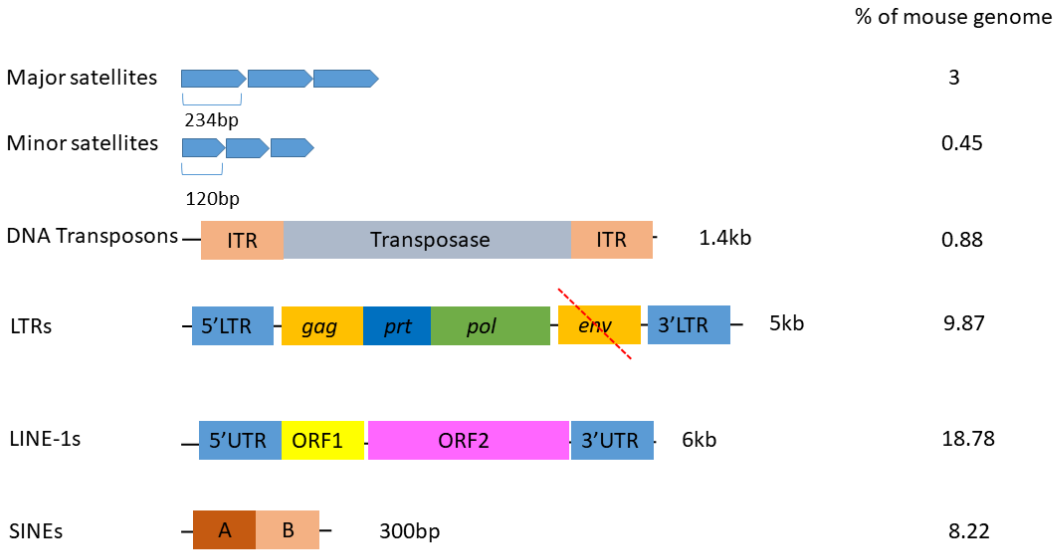


Figure 1.3. Repetitive elements in the murine genome. Structure of the main repeat classes in the mouse genome with estimates of their abundancy as a percentage of the mouse genome (Waterston et al., 2002). The length in bp of a full length i.e. intact repeat unit is displayed. A red dashed line through the *env* (envelope) gene of the LTR indicates that some LTR retrotransposons e.g. IAPs have lost this gene. ITR (Inverted terminal repeat), gag (group specific antigen), prt (protease), pol (polymerase), LTR (long terminal repeat). SINEs contain A and B box promoter sequences of RNA Pol iii. Concept for this figure was taken from (Martens et al., 2005)

The consequences of retrotransposition are potentially severe as new insertions could result in insertional muatagenesis or genomic instability. 10-12% of spontaneous mutations in mice are the result of retrotransposition (Maksakova et al., 2006). In humans there are over 120 identified retrotransposition events resulting in disease (Hancks and Kazazian, 2016). Furthermore retrotransposons could alter the expression of neighbouring genes for example by causing aberrant splicing, by recruiting silencing complexes or acting as alternative promoters or enhancers for host genes as many LTRs contain TF binding sites (Garcia-Perez et al., 2016). Furthermore, proteins expressed by the retrotransposon may impact on cell function. Therefore suppression of these events is vital.

Transcriptional repression through DNA methylation plays a major role in this; indeed the bulk of methylation is found at repetitive sites, and it has been proposed that DNA methylation is present in mammals specifically for the purpose of suppressing these elements (Zemach and Zilberman, 2010).

It is also recognised that retrotransposition events have had evolutionary advantages in the past. Indeed the co-option of transposable element sequences by the host genome has allowed the advantageous expression of genes or different gene isoforms in particular tissues (Cordaux and Batzer, 2009). An important example of this is in the immune system where certain ERVs in human and mouse are found to be enriched near interferon

responsive genes. These ERVs contain binding sites for the TF STAT1, which plays a role in mediating the interferon response, and drive expression of downstream immune genes. The deletion of one such ERV in human cell lines impairs the antiviral response (Chuong et al., 2016). They therefore can act as a vital resource for evolution and expansion of the transcriptome.

It was previously believed that retrotransposition events only occurred in the early developing embryo, germ cells and tumour tissue (Adams, 2017; Hancks and Kazazian, 2012; Kano et al., 2009; Shukla et al., 2013; Solyom et al., 2012; van den Hurk et al., 2007). However, in recent years there has been a growing body of evidence for somatic retrotransposition of LINEs during neurogenesis.

Firstly, it has been shown that many LINE1s contain *Sox2* and *TCF/LEF* (important in *Wnt* signalling) TF binding sites within their promoter. In neural stem cells, *Sox2* can bind to the 5'UTR of LINE1 and repress its promoter activity. During neural differentiation, however, *Sox2* is down-regulated, resulting in de-repression. This is coupled with up-regulation of activating *Wnt* signalling, resulting in expression of LINE1s (Kuwabara et al., 2009; Muotri et al., 2005).

Retrotransposition assays have demonstrated the ability of an engineered LINE1 reporter construct to fully retrotranspose in human tissue culture models of neural development (Coufal et al., 2009) and in rodent brains (Muotri et al., 2005) although this gives no information regarding endogenous retrotransposition. Mosaicism found within the human brain, however, would support somatic retrotransposition events, with brain specific LINE1 insertions being found in human post-mortem tissue (Baillie et al., 2011; Evrony et al., 2012; Upton K et al., 2015). There are conflicting reports on the frequency of these retrotransposition events due to different methodologies used: with estimates ranging from 0.07 (Evrony et al., 2012) to 13.7 somatic insertions per neuron (Upton K et al., 2015). Furthermore, post-mitotic human neurons have also been shown to support retrotransposition of an engineered LINE1 vector *in vitro* (Macia et al., 2017).

This may be an important source of generation of neural diversity within the brain. On the other hand, retrotransposon deregulation has been associated with neurological diseases including multiple sclerosis and several psychiatric disorders. Studies reporting such associations include those which identify an insertion that may cause a disease and those reporting an increase in mRNA levels or retrotransposon copy numbers (reviewed in (Bodea et al., 2018)).

1.3.2 Distribution of DNA methylation

DNA methylation is not evenly dispersed throughout the genome, but instead shows preferential distribution at certain sites dependent on underlying DNA sequence, and performs different roles depending on its genomic context.

1.3.2.1 CpG Islands and promoters

As mentioned earlier around 70% of CpG sites are methylated, however, there exists within the genome ~1kb stretches with higher than expected CpG content termed “CpG islands” (CGIs) which break this observation as they are largely unmethylated (Bird et al., 1985). Given their unmethylated state, the majority (~80%) of CGIs are generally transcriptionally permissive (Tazi and Bird, 1990) it is therefore unsurprising that ~70% of annotated gene promoters are associated with a CGI (Saxonov et al., 2006).

CpG dinucleotides occur less frequently than would be expected in the genome given the abundancy of cytosine and guanine bases. This is thought to be due to the spontaneous deamination of methylated cytosine to yield thymine (Bird, 1980; Coulondre et al., 1978). CGIs at promoter sequences may therefore play important roles in protecting these regulatory sequences from mutation.

Their importance in maintaining a transcriptionally active state is highlighted by the fact that almost all housekeeping gene promoters, which must remain active in all cell types, are associated with a CGI; whereas promoters not associated with CGIs tend to be at genes with tissue specific expression, thereby allowing developmental regulation of these genes in differentiated tissues (Larsen et al., 1992; Zhou et al., 2008). CGIs typically contain the active H3K4me2 mark, which may enable maintenance of a hypomethylated state by preventing binding of the DNA methylation machinery (Guenther et al., 2007). Indeed, *Dnmt3l*, a cofactor for DNA methylation, only recognises and recruits the DNA methylation machinery to unmodified H3K4 (Ooi et al., 2007).

Methylation at promoters as a whole is low, but is correlated with CpG density. These regions have been divided into high-density, intermediate-density and low-density CpG promoters (HCPs, ICPs and LCPs). LCPs are more likely to be highly methylated, but this does not correlate with transcriptional repression. In contrast HCPs are usually unmethylated, but on the occasion that they are methylated the nearby gene is repressed. ICPs, although generally unmethylated in pluripotent cells, are more likely to become methylated as development proceeds and this is associated with gene silencing (Meissner et al., 2008b; Weber et al., 2007).

1.3.2.2 Gene body methylation

In contrast to CGIs, the remainder of the mammalian genome is extensively methylated - in a large part to repress transposable elements in somatic cells. Gene bodies also demonstrate this extensive methylation, although the function of this is less well understood.

Methylation at gene bodies actually appears to be a feature of transcriptionally active genes (Wolf et al., 1984). One proposed role is that this prevents spurious transcription from occurring outside of the transcription start site (TSS) from non-canonical promoters within gene bodies (Neri et al., 2017). Another proposed function is in control of splicing. Exons are more highly methylated than introns and constitutively spliced exons contain higher levels of DNA methylation than do alternative exons (Choi, 2010; Gelfman et al., 2013). DNA methylation at exons may recruit MECP2 and HDACS causing slowing of Pol II elongation resulting in exon inclusion (Maunakea et al., 2013).

1.3.2.3 Non CpG Methylation

In mammalian genomes, methylation was previously thought to occur only in the CpG context: even though non-CpG methylation is widespread in plant genomes (Feng et al., 2010b). This was until advanced sequencing techniques allowed the discovery of non-CpG methylation in mouse and human embryonic stem cells, in which it accounts for almost 25% of methylated cytosines (Lister et al., 2011; Ramsahoye et al., 2000; Ziller et al., 2011). This non-CpG methylation then largely disappears upon differentiation, is almost absent from somatic cells, and is regained in induced pluripotent cells (Laurent et al., 2010; Lister et al., 2011; Ziller et al., 2011). CpA was found to be the most abundant non-CpG methylated site and this was the site that lost most methylation upon differentiation, suggesting a specific role in pluripotency (Laurent et al., 2010).

In most somatic cells non-CpG methylation is lost. Neurons are an exception to this, as it is found to be particularly prevalent in adult neural tissue (Guo et al., 2014; Lister et al., 2013). Whole genome bisulphite sequencing on human and mouse fetal and adult frontal cortex revealed negligible non-CpG methylation in fetal brain tissue, but accumulation of this mark in the postnatal brain. The greatest rate of accumulation is seen during the developmentally important processes of synaptogenesis and pruning, disruption of which is thought to be an aetiological factor for psychiatric disorders (Uhlhaas and Singer, 2011). Maximum levels of 1.3 - 1.5% genome wide are reached at the end of adolescence, at which point non-CpG methylation is actually the dominant form of methylation in neurons at 53% vs 47% CpG methylation, given that non-CpG sites are more abundant than CpGs (Lister et al., 2013).

This *de novo* methylation during neuronal maturation correlates with the ongoing high expression of the *de novo* DNA methyltransferase, *Dnmt3a*, in neurons (Feng et al., 2005; Watanabe et al., 2006). Indeed, Guo et al demonstrated that *Dnmt3a* is required for neuronal

non-CpG methylation, as siRNA knock down of *Dnmt3a* in the adult mouse dentate gyrus resulted in depletion of this mark (Guo et al., 2014).

Sites of non-CpG methylation in the brain are largely conserved between individuals and to a lesser extent between mice and humans (Guo et al., 2014; Lister et al., 2013). Whether this is consistent with a functional role or due to conserved chromatin structure is unclear. These findings broaden the potential role of DNA methylation in the brain, as opposed to other somatic tissues, suggesting important roles in maturation of neural circuits into adult life.

1.3.3 5-hydroxmethylcytosine

DNA methylation can be a reversible process. Demethylation can occur in a passive manner upon cell division simply by the lack of methylation maintenance on daughter strands. However, it has long been recognised that an active demethylation process must also exist. This was recognised by the fact that, in early development, the paternal genome in the mouse zygote undergoes rapid demethylation prior to any DNA replication taking place (Mayer et al., 2000; Oswald et al., 2000; Santos et al., 2002), although this is not a feature of zygotic development in all mammals (Beaujean et al., 2004a; Beaujean et al., 2004b; Young and Beaujean, 2004). In addition, outside an embryonic context, DNA demethylation has been demonstrated in post-mitotic cells (Klug et al., 2010).

The mechanisms behind this active demethylation remained elusive until relatively recently. The conundrum being that direct removal of the methyl group is a thermodynamically unfavourable reaction, with no evidence that this could occur under physiological conditions. However, the rediscovery of 5-hydroxymethylcytosine (5hmC), an oxidised form of 5mC (Penn et al., 1972), and the ten-eleven translocation (TET) enzymes shed new light on potential demethylation pathways. The TET enzymes belong to the TET/J-binding family of α-ketoglutarate iron-dependent dioxygenases. In mammals, 3 have been described to date (TET1-3). *In vitro* work demonstrated that TET enzymes could oxidise 5mC containing oligonucleotide substrates to 5hmc and then further oxidise this to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2010; Ito et al., 2011; Tahiliani et al., 2009). These oxidised bases are then recognised and removed by thymine DNA glycosylase (TDG) and replaced by unmodified cytosine bases as part of DNA repair pathway activity (Cortazar et al., 2011; He et al., 2011).

Far from simply being a demethylation intermediate, 5hmC is considered an epigenetic mark in its own right (Thomson and Meehan, 2017). The mark is several hundred fold more abundant than 5fC and 5caC (Ito et al., 2011), suggesting that a small subset of 5hmC marks are committed to demethylation but the majority may be playing a different role, although it could be the case that 5fC and 5caC are turned over more rapidly.

Unlike 5mC, which is relatively evenly distributed between different tissues, 5hmC levels vary and are particularly abundant in the brain and accumulate during development (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Nestor et al., 2012; Szulwach et al., 2011). The mark is found to be enriched over actively transcribed gene bodies related to neural function and is largely absent from non-genic regions (Hahn et al., 2013; Jin et al., 2011; Lister et al., 2013; Mellen et al., 2012). 5hmC is enriched in genes relating to synaptic plasticity, which need to continually change their expression in the context of activity: therefore demethylation mechanisms may play a key role in the accumulation of 5hmC (Khare et al., 2012).

1.4 Mechanisms of DNA methylation

The deposition of 5mC occurs in two different contexts; that of maintenance and that of *de novo* methylation. The requirement to inherit methylation marks upon cell division relies on maintenance methylation, where methyl marks on the parent strand are faithfully replicated on the daughter strand during replication. On the other hand, *de novo* methylation refers to the acquisition of novel 5mC sites as would occur during differentiation.

The enzymes responsible for catalysing the methylation of cytosine are the DNA Methyltransferases (DNMTs), which are highly conserved between species (Kumar et al., 1994). In mammals, the main DNMTs are the maintenance *Dnmt1* and the *de novo* methyltransferases *Dnmt3a and Dnmt3b* (Barau et al., 2016; Goll and Bestor, 2005; Okano et al., 1999; Yoder et al., 1997). There is, however, evidence for functional overlap between the DNMTs.

1.4.1 The DNA methyltransferases

1.4.1.1 *Dnmt1*

Maintenance of DNA methylation is achieved primarily through *Dnmt1.* The *Dnmt1* enzyme shows a 5-30 fold preference for binding hemi-methylated substrates - such as would be found during DNA replication with the parent strand retaining methyl marks, but the nascent strand lacking these (Goyal et al., 2006; Yoder et al., 1997). As one might expect, *Dnmt1* is therefore highly expressed in dividing cells and is localised to replication foci during S phase (Goll and Bestor, 2005; Leonhardt et al., 1992). Despite being called the maintenance methyltransferase, *Dnmt1* may also have roles in *de novo* methylation. For example, it has

been shown to have greater de novo methylation activity on unmethylated substrates than the *Dnmt3s in vitro* (Jair et al., 2006).

Dnmt1$^{-/-}$ mouse ES cells are viable despite having a severely hypomethylated genome but undergo apoptosis when induced to differentiate (Fouse et al., 2008; Lei et al., 1996; Li et al., 1993; Panning and Jaenisch, 1996). Its importance during differentiation and development is highlighted by the fact that knockout mice are embryonically lethal by E8.5 showing global hypomethylation, retrotransposon activation and misexpression of imprinted and X chromosome genes (Lei et al., 1996; Walsh et al., 1998). However, nuclear transfer experiments demonstrated that hypomethylated nuclei from ES cells null for all three *Dnmts* could support mouse pre-implantation development right up to the blastocyst stage in an equivalent fashion to WT ES nuclei. *Dnmt* null cells also contributed to normally hypomethylated extra-embryonic tissues. These observations suggest that extraembryonic-lineage cells, ES cells and zygotes during pre-implantation development, can survive and proliferate in the absence of DNA methylation (Sakaue et al., 2010).

1.4.1.2 *Dnmt3s*

The DNMT3s are primarily responsible for *de novo* DNA methylation allowing the establishment of new methylation patterns during development. They were discovered after the generation of *Dnmt1* null ES cells demonstrated retention of a small amount of DNA methylation (Lei et al., 1996). These enzymes are highly expressed in ES cells, but are down-regulated after differentiation and expressed at low levels in somatic cells, with some exceptions, including in the nervous system (Feng et al., 2005; Goto et al., 1994; Okano et al., 1998).

These enzymes are highly homologous in their catalytic C-terminal domains, but show subtle differences in their regulatory N terminal domains, which may account for their non-overlapping target specificity (Chen et al., 2004; Dhayalan et al., 2010; Ge et al., 2004; Otani et al., 2009; Zhang et al., 2010). For example *Dnmt3a,* but not *Dnmt3b,* is required for imprinting in germ cells (Kaneda et al., 2004) and *Dnmt3b* is required for methylation of pericentromeric repeats - a function that cannot be replaced by *Dnmt3a* (Okano et al., 1999). There does, however, seem to be some redundancy between the *Dnmt3s* as only double knockout, and not single knockout, ES cells failed to methylate injected retroviral vectors (Okano et al., 1999).

In keeping with their different roles, *Dnmt3a*$^{-/-}$ and *Dnmt3b*$^{-/-}$ mice show different phenotypes. *Dnmt3a*$^{-/-}$ mice appear normal at birth but become runted and die at 4 weeks. Conditional germ cell mutants show hypomethylation at imprinted regions resulting in infertile males and offspring of mutant mothers are lethal by E11.5 (Kaneda et al., 2004). *Dnmt3b* knockout mice on the other hand are embryonic lethal at E14.5-E16.5, with hypomethylation of

satellite repeats, growth impairment and rostral neural tube defects. Double knockout embryos show a similar phenotype to *Dnmt1*$^{-/-}$ mice with development arresting shorty after gastrulation and retrotransposon hypomethylation, although not to the same extent (Okano et al., 1999).

It has been proposed that DNMT3 enzymes are also capable of maintenance methylation, at least *in vitro*, as prolonged culture of *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ ES cells resulted in progressive loss of methylation at both repeats and single copy genes, which was not seen in prolonged culture of wild type or single knockout cells (Chen et al., 2003)(Chen et al., 2003)(Chen et al., 2003)(Chen et al., 2003)(Chen, Ueda et al. 2003). As *Dnmt1*$^{-/-}$ cells demonstrate an immediate 90% loss of DNA methylation, but these double knockout cells required prolonged passage to achieve similar levels of hypomethylation; the authors proposed that while *Dnmt1* is the major maintenance methyltransferase, the *Dnmt3s* may act as "proof-readers" filling in hemi-methylated sites missed by *Dnmt1*.

Dnmt3c was recently discovered in rodents. Previously annotated as a pseudogene, it was found to be exclusively expressed in male germ cells - where it plays a role in methylating and repressing evolutionary young retrotransposons with knockout male mice being sterile (Barau et al., 2016).

1.4.2 Methyl binding proteins

DNA methylation may directly prevent transcription by inhibiting the binding of regulatory proteins or transcription factors, or indirectly result in a more compacted chromatin structure. This occurs as 5mC is recognised or "read" by methyl binding proteins which then recruit co-repressors. There are three main families of methylation readers; the methyl-CpG-binding domain (MBD) family (Hendrich and Bird, 1998; Meehan et al., 1989), the Kaiso-like protein family (Prokhortchouk et al., 2001) and the SET and RING-finger associated (SRA) domain proteins (Johnson et al., 2007).

The MBD family is the best characterised of these, and contains multiple proteins MBD1-6 and MECP1&2 (Brown et al., 2008; Laget et al., 2010; Meehan et al., 1992; Meehan et al., 1989; Ohki et al., 2001; Scarsdale et al., 2011; Screaton et al., 2003). Although all contain methyl binding domains, each protein shows a different affinity for binding 5mC and has different functions (Hendrich and Bird, 1998; Hendrich et al., 1999). In mammals, MBD1, MBD2 and MECP2 can bind to methylated CpG sites resulting in transcriptional repression through the recruitment of co-repressor complexes, which vary depending on the MBD protein in question but include histone deacetylases, lysine methyltransferases and chromatin remodellers (Jones et al., 1998; Nan et al., 1997; Nan et al., 1998; Ng et al., 2000).

Despite the extensive evidence for the role of MBDs in gene repression, it should be borne in mind that they are not essential for development in mouse knockout models with most only having mild phenotypes, at least in comparison to *Dnmt* knockouts, and limited de-repression of genes (Guy et al., 2001; Hendrich et al., 2001; Martin Caballero et al., 2009; Prokhortchouk et al., 2006; Sasai and Defossez, 2009). This could be due to some redundancy between the proteins, or that DNA methylation without recruitment of methyl binding proteins is sufficient to maintain repression.

1.4.3 Cofactors of DNA Methylation

There are several proteins which have been shown to be of import in facilitating DNA methylation. These cofactors can exert their influence by recruiting DNMTs, stimulating their activity or by modifying the local chromatin structure to alter the accessibility of the DNA. In this section, I will discuss known cofactors which will lead onto a discussion about the protein of interest in this dissertation, *Lsh.*

1.4.3.1 *Dnmt3l*

Another mammalian specific member of the *Dnmt* family is *Dnmt3l*. This protein shares some sequence homology to the DNMT3s, however is truncated and missing critical amino acid motifs associated with catalytic activity (Aapola et al., 2000).

On its own, *Dnmt3*l has not been shown to have any methyltransferase activity - however it can increase the catalytic activity of *Dnmt3a* and *Dnmt3b* by around 15 fold by directly binding to the catalytic domains and inducing a conformational change (Gowher et al., 2005). That it plays a role in *de novo* methylation is supported by its expression pattern being high in the early embryo and becoming down-regulated in somatic cells. It is also expressed at particular stages during germ cell development that coincide with peaks of *de novo* methylation (2011; Bourc'his et al., 2001; Sakai et al., 2004).

Knockout mice are viable and survive to adulthood - but males are azoospermic, which may be due to the inappropriate activation of retrotransposons during spermatogenesis. Conversely, female knockout mice are fertile but their heterozygous offspring die in mid-gestation. This is due to the fact that oocytes from knockout females showed grossly reduced methylation at ICRs, with embryos having biallelic expression of imprinted genes (Bourc'his et al., 2001; Hata et al., 2002). This closely resembles the phenotype of targeted knockout of *Dnmt3a* in germ cells (Kaneda et al., 2004).

1.4.3.2 *Uhrf1*

An essential cofactor for *Dnmt1* is *Uhrf1.* This protein contains a SET and ring-associated domain that binds to hemi-methylated CpGs (Achour et al., 2008; Arita et al., 2008;

Avvakumov et al., 2008) and co-localises with *Dnmt1* during S phase at DNA replication sites thereby demonstrating its role in targeting *Dnmt1* to hemi-methylated replicating DNA (Bostick et al., 2007; Sharif et al., 2007). *Uhrf1*$^{-/-}$ embryos show a very similar phenotype to *Dnmt1*$^{-/-}$, with global hypomethylation and loss of repression at retrotransposons and imprinted regions and arrest in development shortly following gastrulation (Sharif et al., 2007).

1.5 *Lsh*

Of particular interest is another DNA methylation cofactor *lymphoid specific helicase (Lsh)* so named because it was initially identified as a transcript highly expressed in lymphocytes (Geiman et al., 1998; Jarvis et al., 1996). In fact it is highly expressed in proliferating cells, therefore postnatally there is preferential expression in the thymus, bone marrow, spleen and testis however it is ubiquitously expressed during embryonic development, including being highly expressed in the developing mouse embryonic brain (Raabe et al., 2001).

Its potential role in DNA methylation was implied following the discovery that a mutated version of its homolog in *Arabidopsis thaliana,* named *Ddm1 (Decrease in DNA Methylation 1),* resulted in a 70% reduction of methylation (Vongs et al., 1993). Subsequently *Lsh*$^{-/-}$ mice were shown to have a 30-40% reduction in global DNA methylation levels predominantly at satellite repeats (Dennis et al., 2001; Geiman et al., 2001; Tao et al., 2011).

1.5.1 *Lsh* structure and classification

Lsh may also be referred to as *Smarca6* (SWI/SNF related matrix associated actin-dependent regulator of chromatin subfamily a member 6), *Pasg* (proliferation associated SNF2-like gene) and *Hells* (helicase, lymphoid specific). In this thesis *Lsh* will refer to the murine homolog and *Hells* to the human homolog.

In mice, the gene is found on chromosome 19 at C3-D1 and in humans at position 10q23.33 (Geiman et al., 1998). The *Lsh* gene contains 22 coding exons encoding important conserved functional domains including the ATP binding domain and the DEAH box vital for its function (Ren et al., 2015; UniProt Consortium, 2018). *Lsh* is highly conserved in mammals, showing a 91% overall sequence homology and 95% sequence homology within ATPase domains (Meehan et al., 2001)(Figure 1.4).

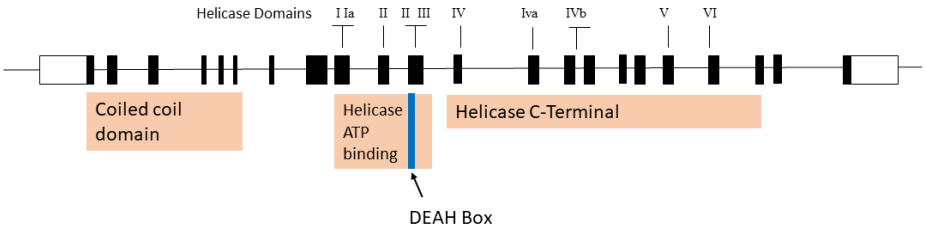


Figure 1.4. The murine *Lsh* gene. Exons are displayed in black with the conserved helicase domains and structural domains highlighted.

Due to homology in sequence across the ATPase domain and helicase motifs, *Lsh* is classified as a member of the SNF2-like family of chromatin remodellers (Geiman et al., 1998). Members of this family utilise energy derived from ATP hydrolysis to alter chromatin structure at the level of the nucleosome. SNF2 ATPases are usually found in large multi-subunit complexes. However, to date, *Lsh* has not been found to act within a large complex (Myant and Stancheva, 2008). It is an exclusively nuclear protein and accumulates at pericentromeric heterochromatin, in keeping with its important role in methylating satellite sequences found at these sites (Yan et al., 2003).

As previously stated, it is highly expressed in proliferating cells: therefore is universally expressed in the early embryo. However, expression becomes more restricted as development proceeds. High expression is seen in the brain at E12.5-E14.5 but expression in adult tissue becomes restricted to those with proliferative activity including testis, small intestine, thymus, bone marrow and spleen (Raabe et al., 2001).

1.5.2 *Lsh-/-* Mice

The absolute requirement for *Lsh* for normal murine development is demonstrated in a non-viable knockout mouse generated by Geiman *et al.* This mouse carries a deletion across helicase domains I, Ia and part of domain II thereby spanning the ATPase domain and DEAH box. Knockout pups survive to term but die within a few hours of birth. No pathology is found in the knockout embryos before gestational day 14. However, new born pups exhibit growth retardation and severe renal lesions. Histology of many other organs including the brain revealed no abnormalities (Geiman et al., 2001).

Sun et al generated an alternative knockout mouse model carrying a deletion across exons 10-12 of the *Lsh* gene containing helicase domains II, III and IV. This resulted in a hypomorphic mutation encoding a truncated form of the LSH protein expressed at low levels. 60% of these mice die shortly after birth, although 40% survive up to a few weeks following birth. Growth retardation is noted in these mice from E12.5 and becomes more severe as development proceeds. At birth, histopathological examination revealed diffuse alveolar

atelectasis but other organs examined, including the brain and kidney, showed no abnormalities. In those pups that survived the postnatal period, an early aging phenotype was observed with development of grey hair, alopecia, cachexia, unstable gait and kyphosis. Histopathological examination at 2 weeks of age revealed lymphoid depletion in the thymus and renal degeneration (Sun et al., 2004). *Lsh* is therefore essential for murine survival.

1.5.3 *Lsh* and DNA methylation

Similar to *Ddm1* mutant plants, *Lsh*$^{-/-}$ mice show defects in DNA methylation at repetitive elements. Sun *et al* demonstrated a 30% reduction in global methylation in E13.5 hypomorphic embryos and a specific hypomethylation at minor satellite and Intracisternal A Particles (IAPs), a type of LTR, with derepression of IAPs in tissue from 15 day old mice (Sun et al., 2004). Similarly, repeat hypomethylation is seen in tissues derived from the Geiman knockout mouse model described previously. This was demonstrated by Southern analysis of satellite, SINEs, IAPs and LINE1s repeats. In this study hypomethylation of a small number (3) of selected single copy genes and the ICR of the *H19* gene was also demonstrated (Dennis et al., 2001).

Other studies have also pointed towards *Lsh* playing a role in methylation at single copy genes. This includes a specific role in methylating stem cell genes - as demonstrated by siRNA knockdown of *Lsh* in mESCs, in which promoter DNA methylation and subsequent repression of pluripotency associated genes upon differentiation was impaired. Inappropriate promoter hypomethylation and expression of *Oct4* was also seen in E8.5 embryos but by E18.5 *Oct4* was successfully repressed despite continued hypomethylation (Xi et al., 2009).

Other developmentally important genes have been reported to be inappropriately regulated in the absence of *Lsh*. These include promoter hypomethylation and biallelic expression of the imprinted *Cdkn1c* gene in *Lsh*$^{-/-}$ embryos (Fan et al., 2005b). *Hox* genes are clusters of genes which are silenced as differentiation proceeds. However, these genes remain inappropriately expressed in *Lsh*$^{-/-}$ MEFs and tissues derived from *Lsh*$^{-/-}$ embryos at day 18 of gestation - which authors propose is due, at least in part, to promoter hypomethylation (Xi et al., 2007).

The disadvantage in these studies is the investigation of methylation and transcription at only selected loci. The use of genome-wide sequencing has revealed further information into the role of *Lsh*. Whole genome bisulphite sequencing by Yu *et al* in *Lsh*$^{-/-}$ MEFs confirmed ~60% hypomethylation at repeat elements with a subset of these, in particular ERVs and satellites, being derepressed as shown by RNA-Seq. This form of analysis also revealed large chromosomal domains which were hypomethylated in the absence of *Lsh* and that these regions frequently overlapped with lamin B1 attachment domains (LADs), gene poor repeat rich regions of the genome. *Lsh* may therefore be targeted to nuclear compartments rather

than to underlying sequences. Sequencing also revealed that the majority of promoter regions in KO cells have reduced methylation, particularly those of low CpG content, although only 2% of genes showed differential expression (Yu et al., 2014b). Providing further evidence for *Lsh* influencing methylation over discrete chromosomal domains, Tao *et al* produceded a methylation map of *Lsh*$^{-/-}$ MEFs using methylated DNA immunoprecipitation (MeDIP) combined with whole-genome tiling microarray and found the average size of differentially methylated regions to be 50-200kb with well-defined boundaries (Tao et al., 2011).

Myant *et al* found many (1246) genes were differentially expressed in *Lsh*$^{-/-}$ MEFs - including imprinted and Hox genes. However, only 10% of these were associated with a change in promoter methylation. This study used methyl-CpG binding domain affinity purification (MAP) combined with promoter-specific microarrays, to demonstrate hypomethylation at ~20% of promoter regions. This also revealed hypermethylation at 5% of promoters tested indicating a rearrangement of DNA methylation patterns at other loci throughout the genome (Myant et al., 2011).

At the beginning of this project it was unknown when *Lsh* contributed to DNA methylation. Then Ren *et al* derived mES cells from *Lsh*$^{-/-}$ blastocysts, induced them to differentiate into neuroepithelial lineage, and examined DNA methylation at repeat sequences by bisulphite sequencing. This revealed that, despite the fact that *Lsh* is highly expressed in ES cells and down-regulated upon differentiation, it was required for methylation upon differentiation and reintroduction into ES cells did not restore DNA methylation until the cells were differentiated. Upon retinoic acid induced differentiation KO cells displayed hypomethylation at IAP sequences (86% versus 49%), minor satellite sequences (79% versus 28%), and LINE1 elements (80% versus 55%). This methylation defect was rescued by re-expression of WT *Lsh*, but not *Lsh* carrying a mutation in the ATP or DEAH domain. This beautifully demonstrated that *Lsh* was not required to maintain DNA methylation in ES cells, but contributes to DNA methylation in the context of differentiation (Ren et al., 2015).

1.5.4 *Lsh* and *de novo* methylation

Lsh is predominantly thought to contribute to *de novo* methylation. This conclusion has been drawn from several bodies of evidence. Firstly by transfecting episomal vectors into *Lsh*$^{-/-}$ MEFs it was demonstrated that *Lsh* was required to establish novel methylation patterns but was not required for maintenance of previously methylated vectors (Zhu et al., 2006). Secondly, in the study by Ren *et al* mentioned above, methylation was gained at repeat sequences upon differentiation in the absence of *Lsh;* just not to the same extent as in WT cells suggesting deficits in the de novo methylation pathway (Ren et al., 2015). Finally co-immunoprecipitation studies have revealed interactions between *Lsh* and the de novo methyltransferases *Dnmt3a* and *Dnmt3b* (Myant and Stancheva, 2008; Zhu et al., 2006).

However, one cannot completely exclude a role for *Lsh* in maintenance methylation as knockdown of *Lsh* in primary human fibroblasts leads to progressive loss of methylation at repeat sequences (Suzuki et al., 2008) and it has also been shown to directly interact with the maintenance methyltransferase *Dnmt1* (Dunican et al., 2015).

1.5.5 Mechanisms of action

How *Lsh* contributes to DNA methylation patterns and transcriptional repression is less clear. It contains no known methyltransferase domain, therefore must exert its action indirectly. There is evidence for its action as a chromatin remodeller similar to *Ddm1* in *Arabidopsis*, also there is evidence for it acting as a recruiting protein for DNMTs. Finally, it has been shown to act in concert with other histone modifications, which may together influence local chromatin structure. Evidence for these mechanisms of action will be discussed below.

1.5.5.1 Recruitment of *Dnmts*

The association between *Lsh* and *Dnmts* has been well demonstrated, as it has been shown to co-immunoprecipitate with these enzymes (Dunican et al., 2015; Myant and Stancheva, 2008; Xi et al., 2009; Zhu et al., 2006). In keeping with its proposed role in *de novo* methylation most evidence is for an association with *Dnmt3b* activity. Association of *Dnmt3b* at target loci such as IAP elements and *Hox* genes is lost in the absence of *Lsh* (Ren et al., 2015) (Xi et al., 2009; Xi et al., 2007). Furthermore the ability of *Lsh* to repress a reporter gene was impaired on knockout of DNMTs but interestingly the catalytic activity of DNMTS was not required for repression (Myant and Stancheva, 2008). If this was the major mechanism by which *Lsh* influences DNA methylation, one may expect to see similarity between $Lsh^{-/-}$ and $Dnmt3b^{-/-}$ hypomethylated regions. Indeed this is the case as investigated by Dunican *et al* in MEFs. However, this paper also reported that compared to $Dnmt3b^{-/-}$ MEFs, $Lsh^{-/-}$ MEFs had more severe hypomethylation defects in LTRs and greater expression of hypomethylated satellite and IAP sequences (Dunican et al., 2013). It would therefore appear that *Lsh* acts primarily in concert with *Dnmt3b*, however does have a role independent of this enzyme at certain loci.

Lsh may therefore act as a scaffolding protein recruiting DNMTs to target sites. However Ren *et al* demonstrated that the ATPase activity of *Lsh* is essential for DNA methlation and DNMT3B recruitment to repeat sequences (Ren et al., 2015) pointing to a primary action in chromatin remodelling as will be discussed in the next section.

1.5.5.2 Chromatin remodelling

As previously stated, *Lsh* is classed as a member of the SNF2-like family of chromatin remodellers. Its homolog *Ddm1* was shown several years ago to directly induce nucleosome

remodelling (Brzeski and Jerzmanowski, 2003). However, demonstration of this activity in *Lsh* remained elusive until recently. Indeed Jenness *et al* found purified HELLS possessed little remodelling activity in a restriction enzyme accessibility assay, whereby remodelling exposes a restriction site within the nucleosome. However, addition of CDCA7, which is incapable of remodelling action on its own, resulted in robust remodelling activity. These proteins were shown to form a stoichiometric complex on chromatin and act as a bipartite nucleosome remodelling complex the authors termed the CDCA7-HELLS ICF-related nucleosome remodelling complex (CHIRCC) (Jenness et al., 2018).

This remodelling has been proposed to allow access for the DNMTs to the DNA (Meehan et al., 2001). Indeed nucleosomal DNA cannot be efficiently methylated by *Dnmts* unless in the presence of chromatin remodellers (Felle et al., 2011; Schrader et al., 2015; Takeshima et al., 2006). *Ddm1* has been suggested to promote accessibility at H1 dense repressed heterochromatin, improving deposition of DNA methylation at these sites (Zemach et al., 2013). Evidence for *Lsh* acting as a nucleosome remodeller *in vivo* came from the aforementioned study by Ren *et al* who differentiated *Lsh*$^{-/-}$ ES cells. In order to create a simultaneous map of nucleosome footprint and DNA methylation they carried out nucleosome occupancy and methylome sequencing assay (NOMe Seq). This reported high nucleosome occupancy at IAP and LINE1 elements in both KO and WT ES cells, but reduced occupancy in KO cells following differentiation at 59% vs 98% at IAPs and 59% vs 85% at LINE1s, which also demonstrated hypomethylation. Nucleosome occupancy was restored by reintroduction of WT *Lsh* but not ATP or DEAH domain mutants (Ren et al., 2015). This indicates a requirement for *Lsh* for nucleosome density at these repeat sites, however does not necessarily indicate that this nucleosome occupancy is a direct consequence of *Lsh* remodelling, nor that it occurs prior to DNA methylation.

This same group more recently attempted to further the understanding of the dynamics involved by creating a tethering system recruiting a *Gal4-Lsh* fusion protein to an engineered *Oct4* locus and investigated a range of epigenetic modifications at this site. In ES cells association of this protein to the tethered site reduced chromatin accessibility and active histone modifications, whilst having no effect on expression (as assessed by GFP reporter gene expression) or DNA methylation. When these cells were induced to differentiate there was an increase in repressive histone modifications, gain of DNA methylation, and transcriptional repression of the reporter gene (Ren et al., 2018). This provides evidence for chromatin remodelling occurring prior to *de novo* DNA methylation pointing to a priming role for *Lsh* in ES cells whereby local changes in chromatin structure at this stage prime the gene for later DNA methylation and repression. This may explain why *Lsh* is expressed so highly in ES cells despite the lack of evidence for it playing a great contribution to DNA methylation at this stage.

1.5.5.3 *Lsh* and histone modifications

Lsh has been proposed to influence the epigenome through mechanisms other than DNA methylation, which will be discussed here.

A wide-spread disturbance of histone modifications have been observed in *Lsh*$^{-/-}$ models. Dunican *et al* demonstrated changes in the histone code over repeats, correlating with transcription with a decrease in repressive H3K9me3 at de-repressed IAPs and satellites and gains of active H3K4me3. The gains in active histone modifications were also seen over LINE1s, which did not demonstrate transcriptional de-repression, but loss of H3K9me3 was not detected at these sites. Accumulation of repressive H3K27me3 was also seen over repeats (Dunican et al., 2013).

H3K27me3 is deposited by Polycomb group proteins, and other authors have associated these proteins with *Lsh*. Xi *et al* immunoprecipitated components of the Polycomb repressive complex 1 (PRC1) and found *Lsh* associates with these. Furthermore, in *Lsh*$^{-/-}$ MEFs there was a reduction in the PRC1 mark H2A-K116 ubiquitination and the PRC2 mark H3K27me3 across target *Hox* loci (Xi et al., 2007).

Lsh may maintain a repressive state via recruitment of HDACs, as it has been shown to immunoprecipitate with these enzymes and histone deacetylase activity has also been shown to be essential for *Lsh* mediated repression at the *p16* promoter (Zhou et al., 2009).

It has also been proposed to co-operate with the G9a/Glp complex responsible for most of the dimethylation at H3K9 (Tachibana et al., 2008) as there was failure of this complex to associate with target loci in *Lsh*$^{-/-}$ MEFs, with a subsequent reduction in H3K9me2 at these sites (Myant et al., 2011).

Regions differentially enriched for H3K4me1, a mark of putative enhancers, have also been noted in *Lsh*$^{-/-}$ MEFs. Interestingly these putative novel enhancers were clustered near neuronal lineage genes. When these cells were reprogrammed to iPSCs they maintained these marks and subsequently demonstrated enhanced neurodifferentiation potential (Yu et al., 2014a).

The main role of *Lsh* is therefore in DNA methylation. However, either through direct or indirect means, it also has a more widespread influence over the epigenome with effects on transcription.

1.5.6 Genomic instability

Lsh, most probably through its role in DNA methylation at satellite sequences near the centromere, plays a key role in maintaining genome stability. Fibroblasts derived from *Lsh*$^{-/-}$ mice display replicative senescence with increased centrosome numbers and multipolar spindle formation similar to the phenotype of WT MEFs treated with 5-Aza-2'-deoxycytidine, a demethylating agent (Fan et al., 2003) although it should be noted that one paper demonstrated that this was not the case for fibroblasts derived from patients with mutations in *Hells* (Toubiana et al., 2018).

It is also important in maintaining genome stability during meiosis. *Lsh* null oocytes show incomplete synapsis of chromosomes, non-resolution of double strand breaks in asynapsed chromosomes and failure of homologous recombination resulting in severe oocyte loss (De La Fuente et al., 2006).

There is also evidence that *Lsh's* role in maintaining genome stability is independent of its role in DNA methylation. Burrage *et al* treated *Lsh* deficient mouse and human fibroblasts with ionizing radiation and found insufficient repair of DNA double strand breaks. These results were not seen in *Dnmt1*$^{-/-}$ MEFs suggesting they were independent from DNA methylation defects (Burrage et al., 2012). Unoki *et al* also showed evidence for a role for Lsh in non-homologous end-joining (Unoki et al., 2019).

The importance of *Lsh* in DNA methylation and maintenance of genomic stability in humans is reflected by its misregulation in cancer (Fan et al., 2008; Jia et al., 2017; Liu and Tao, 2016; von Eyss et al., 2012; Wang et al., 2015), and by the fact that mutations in this gene are causative for ICF syndrome.

1.5.7 ICF syndrome

Immunodeficiency, Centromeric Instability and Facial Anomalies syndrome (ICF) is a rare autosomal recessive disorder. In 50% of cases it is caused by mutations in *Dnmt3b* and a further 30% are accounted for by mutations in zinc-finger and BTB domain containing 24 (*Zbtb24*) (de Greef et al., 2011). In 2015 Thijssen *et al* also identified causative mutations in cell division cycle associated 7 (*Cdca7*) and *Hells* (Thijssen et al., 2015).

Individuals suffer from recurrent infections which often prove fatal due to immunoglobulin deficiency. They also present with facial dysmorphism which can include hypertelorism, a flat nasal bridge and epicanthus, growth retardation and intellectual disability (Ehrlich et al., 2008). Centromeric instability in mitogen stimulated lymphocytes is a hallmark of this syndrome involving the pericentromeric regions on chromosomes 1, 9, 16 and rarely 2 due to hypomethylation and decondensation at satellite repeats (Jeanpierre et al., 1993; Tuck-

Muller et al., 2000). Treatments involve immunoglobulin infusions and allogenic stem cell transplantation but life expectancy remains poor at 8 years.

Although there is variability in the presentation of ICF syndrome, there are 3 diagnostic criteria which are: immunodeficiency despite the presence of B cells, hypomethylation in satellite DNA and characteristic rearrangements in chromosome 1 and/or 16 in mitogen stimulated lymphocytes (Ehrlich et al., 2006). These rearrangements include pericentromeric breaks, whole arm deletions, multibranched chromosomes, translocations and isochromosomes (Ehrlich, 2003).

The juxtacentromeric hypomethylation is seen across all studied tissues from ICF patients. However, DNA rearrangements are thought mainly to occur in lymphocytes, rather than other tissues - and even in these cells they are thought to occur at a very low rate in vivo (Ehrlich, 2003; Sawyer et al., 1995).

Mouse models have been generated carrying *Dnmt3b* mutations causative for ICF. The majority of these mice display low birth weight and die within 1 day of birth of unknown cause. A very small number survived past weaning age and had facial features reminiscent of ICF syndrome. Newborn pups displayed hypomethylation of satellite repeats and a deficient immune system although unlike in humans this was due to a defect in the T cell lineage (Ueda et al., 2006).

Seven ICF syndrome patients across five families have been identified with causative mutations in *Hells*. The mutations seen in each proband are summarised in Figure 1.5 (Alghamdi et al., 2018; Thijssen et al., 2015).

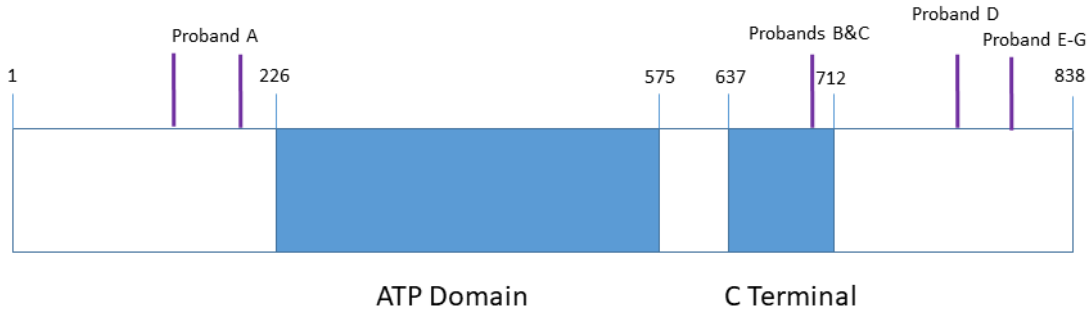


Figure 1.5. Mutations in *Hells* in ICF patients. Schematic representation of Hells protein with mutations identified in 7 patients highlighted. Proband A is a compound heterozygote for a nonsense mutation at Lys204 and a duplication causing insertion of a stop codon at Lys128. Probands B and C were compound heterozygotes with a missense mutation at Gln699 within a conserved helicase domain and an intronic mutation leading to destruction of a splice donor site. Proband D carries a homozygous out-of-frame deletion at Ser762 resulting in a frameshift mutation and introduction of a premature stop codon at exon 20.Probands E-G carry a homozygous in-frame deletion at Leu801. (Adapted from (Thijssen et al., 2015))

The fact that mutations in several different genes can all lead to ICF syndrome has led to the subdivision of this syndrome into: type 1 with causative mutations in *Dnmt3b*, type 2 with causative mutations in *Zbtb24*, type 3 with causative mutations in *Cdca7* and type 4 with causative mutations in *Hells*. Patients with each subtype have been proposed to display differing phenotypes (Velasco et al., 2018a). For example, patients with type 2-4 are said to suffer more from neurological than immunological defects compared to patients with type 1, although the small number of cases and the heterogeneity between individuals limits the extent to which one can draw such conclusions.

A comparative analysis of patients with ICF revealed that each subtype could be distinguished by the methylome. All subtypes contained hypomethylation of pericentromeric repeats and a few common loci, but distinct loci were affected in each subtype. Types 2-4 clustered together and displayed greater hypomethylation at CpG poor regions, with heterochromatic features and at genes implicated in neuronal development. It should be noted that this analysis included only one patient with type 4 ICF (Velasco et al., 2018a).

1.6 Epigenetic dynamics during development

Epigenetic mechanisms are at their most dynamic during embryonic development. Initially this requires a return to a pluripotent state at fertilisation following fusion of cells that have acquired all the epigenetic marks of germ cells. Following this, cells with identical DNA

sequence must rapidly divide and follow different differentiation trajectories to establish different transcriptional profiles. This is achieved through transcription factor networks and wide-scale remodelling of the epigenome.

1.6.1 DNA methylation in the early embryo

During embryonic development extensive demethylation occurs in two separate waves. The first wave occurs in the zygote following fertilisation, and the second in the primordial germ cells (PGC's)(Santos et al., 2002).

The first wave of demethylation is coincident with restoration of pluripotency in the developing zygote. This involves extensive loss of DNA methyl marks, such that by the early blastocyst stage methylation levels are at their lowest. Demethylation however is not complete as methylation at imprinted regions, and some, but not all, repetitive elements are preserved (Brandeis et al., 1993; Lane et al., 2003). The kinetics of demethylation differs between the paternal and maternal genome, occurring much more rapidly in the paternal. This led to the conclusion that the paternal genome underwent an active demethylation process whereas the maternal genome underwent a passive, replication dependent, demethylation (Oswald et al., 2000; Santos et al., 2002). This, however, has been called into question by the presence of oxidised cytosines bases and demethylation beyond what would be expected from replication dependent means in the maternal genome suggesting a degree of active demethylation also (Wang et al., 2014). The importance of this extensive demethylation however is unclear as it does not occur in all mammals (Beaujean et al., 2004a).

The TET enzyme pathway of demethylation described in section 1.3.3 has been proposed as a strong candidate for this active demethylation in zygotes. Evidence for this includes the fact that TETs and 5hmC are enriched in male pronuclei compared to female. Furthermore, *Tet3* deficient zygotes failed to convert 5mC to 5hmC in the paternal genome (Gu et al., 2011). One reason why active demethylation in female pronuclei is far less pronounced may be the higher expression of the PGC7 protein, which associates with H3K9me2 and reduces affinity of TET3 to these sites (Nakamura et al., 2012; Szabo and Pfeifer, 2012).

There is, however, evidence against the TET enzyme pathway being the main mode of active demethylation in the zygote. Importantly the increase in 5hmC levels in male pronuclei does not occur in concert with the reduction in 5mC, but instead is delayed. This was demonstrated by Amouroux et al, who also proposed 5hmC was actually derived from de novo methylation by *Dnmt3a* after they demonstrated a significant decline of 5hmC levels, with no disturbance of 5mC levels in male pronuclei of zygotes treated with a methyltransferase inhibitor (Amouroux et al., 2016). The fact that 5hmC is a stable mark retained through all cleavage stage embryos suggests it may play an important role during

development, and is not just the by-product of 5mC demethylation (Iqbal et al., 2011; Wossidlo et al., 2011).

The second demethylation wave occurs in PGCs at E8.5-E12.5. As PGCs are derived from cells at a later stage of differentiation than the hypomethylated inner cell mass (ICM), they have already begun to accumulate somatic epigenetic features. Therefore, these newly acquired DNA methyl marks need to be removed to restore totipotency and allow laying down of germ-line specific epigenetic patterns. This demethylation goes further than the first wave, as it also has to remove imprinted marks to allow reestablishment in a sex dependent manner (Hajkova et al., 2008).

The initial demethylation event necessitates a subsequent wave of *de novo* methylation in the implanting blastocyst to allow cell fate specification. In this wave 5mC is established at promoters of stem cell specific, germ-line and tissue specific genes dependent on cell fate (Mohn et al., 2008; Ng et al., 2008). This wave continues until tissue specific patterns of DNA methylation are laid down largely by E6.5 (Borgel et al., 2010). *De novo* methylation in PGCs occurs at E14.5 in males and not until postnatally in females (Li et al., 2004; Lucifero et al., 2004) (Figure 1.6). In general DNA methylation is seen as a relatively stable mark that gradually accumulates during a cell's differentiation, progressively restricting cell fate.

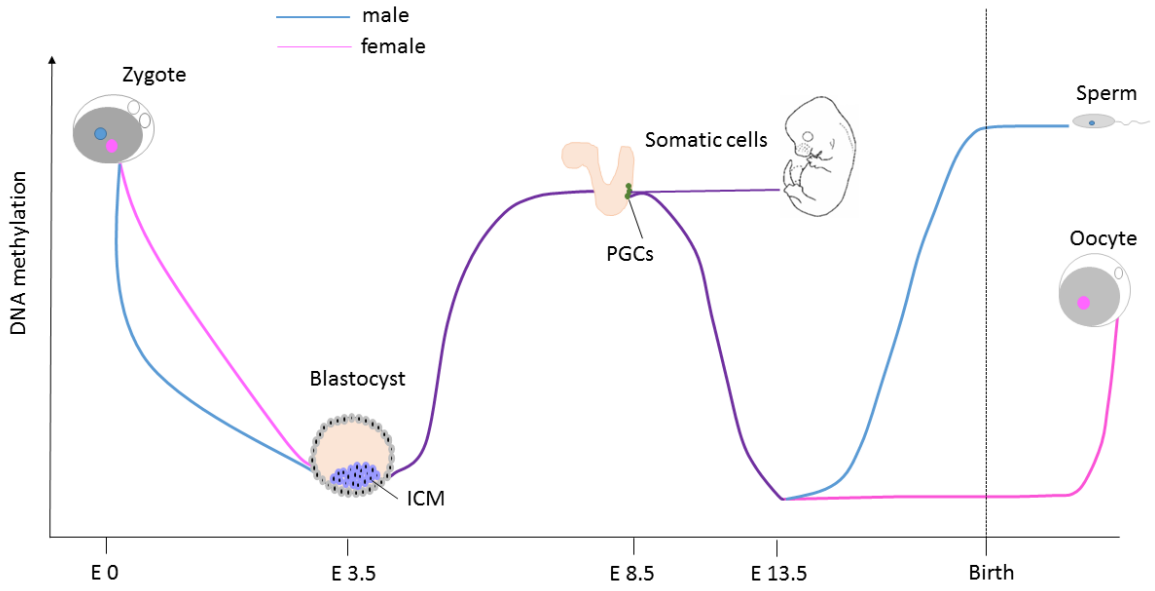


Figure 1.6. DNA methylation during murine embryogenesis. Diagram of DNA methylation patterns during murine embryogenesis. Blue lines represent the paternal genome and pink the maternal. Days of embryonic development are displayed on the x axis. Following fertilisation there is a wave of demethylation resulting in the lowest levels of methylation in the inner cell mass (ICM) of the implanting blastocyst. A wave of *de novo*

methylation then begins and continues until tissue specific patterns of DNA methylation are laid down. At E8.5 demethylation occurs in the migrating primordial germ cells (PGCs). *De novo* methylation then occurs to establish sex specific patterns. This occurs at E13.5 in the male but not until postnatally (P21) in the growing oocyte.

1.6.2 Polycomb during development

As well as profound changes in DNA methylation, histone modifications are also dynamically turned over during development. In section 1.2.3 I alluded to the importance of Polycomb proteins in this process.

Polycomb proteins are highly conserved key chromatin modifiers. They are found in complexes termed Polycomb repressive complex 1(PRC1) and PRC2. PRC1 complexes can be formed of multiple different proteins, however 2 core proteins are present in all; a RING1 protein which mediates ubiquitination of H2AK119 and a Polycomb group ring-finger domain protein. The canonical PRC1 complex also contains a chromobox protein that binds H3K27me3. The functional core of PRC2 contains the histone methyltransferase enhancer of zeste (EZH1 or 2) responsible for depositing the H3K27me3 mark, suppressor of zeste (SUZ12), embryonic ectoderm development (EED) involved in histone deacetylation and histone binding protein RBBP4 or 7 (van der Vlag and Otte, 1999). The counterparts of Polycomb proteins are the Trithorax group proteins which can remove the repressive Polycomb marks and can deposit the transcriptionally permissive mark H3K4me3 (Schuettengruber et al., 2017).

The established model of Polycomb gene repression is that PRC2 is recruited and deposits H3K27me3 providing a binding site for the chromodomain of PRC1, which then catalyses deposition of the repressive mark H2AK119ub1 (Endoh et al., 2012; Nakagawa et al., 2008; Stock et al., 2007; Zhou et al., 2008). However this model cannot be a complete story as the complexes have some non-overlapping targets, non-canonical PRC1 complexes lack a chromodomain and some organisms, whilst having H3K27me3 lack a PRC1 homolog (Gao et al., 2012; Ku et al., 2008; Leeb et al., 2010; Schuettengruber et al., 2007). It has also been proposed that PRC1 can recruit PRC2 via recognition of H2AK119ub1 creating a positive feedback loop perpetuating the mark (Blackledge et al., 2014; Cooper et al., 2014). Furthermore the catalytic activity of RING1B does not appear to be essential for Polycomb function in all contexts (Illingworth et al., 2015). Another means by which the mark could be perpetuated is by influencing higher order chromatin structure. PRC1 complexes can form tightly compacted chromatin structures and bring together distal genomic regions, and this action appears to be independent of ubiquitination (Kundu et al., 2017).

There seems to be some form of cross talk between Polycomb and DNA methylation as they tend to be mutually exclusive, i.e. promoters with high levels of DNA methylation are depleted for H3K27me3 mark and vice versa (Hawkins et al., 2010; Meissner et al., 2008a).

Furthermore, in models of DNA hypomethylation there appears to be redistribution of the H3K27me3 mark to previously methylated regions (Lynch et al., 2012; Reddington et al., 2013).

Polycomb targets are enriched for tissue-specific transcription factors and developmental regulators, such as *Hox* genes, highlighting a role for Polycomb in regulating cell fate during development (Boyer et al., 2006; Lee et al., 2006). In ES cells many important developmental genes are bivalently marked. That is, they are marked by both the active H3K4me3 mark and the repressive H3K27me3 mark. In this situation Polycomb is playing a key role in silencing developmental genes, thereby maintaining pluripotency. This is demonstrated in the spontaneous differentiation of ES cells deficient in *Suz12* or *Eed* which also show de-repression of developmental genes (Boyer et al., 2006; Lee et al., 2006). The necessity for intact Polycomb regulation during development is further highlighted by the embryonic lethality of knockout mouse models (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004).

Repression of developmental genes necessarily has to be transient to allow differentiation to occur at the appropriate developmental time point. Bivalently marked genes are silent but poised for later transcription upon cell fate determination, when they lose their H3K27me3 mark or become more stably repressed, losing H3K4me3 dependent on cell type (Mikkelsen et al., 2007). Poised RNA polymerase II is associated with bivalently marked genes to allow rapid transcription upon differentiation (Brookes et al., 2012). These bivalent regions are not exclusive to ES cells, indeed new bivalent domains arise in neural progenitors for example (Mohn et al., 2008). During neural differentiation of ES cells many more promoters lose and gain H3K27me3 at different stages of neurodifferentiation than change their DNA methylation status (Mohn et al., 2008).

1.6.3 Epigenetic dynamics during neurodevelopment

1.6.3.1 Neurogenesis

The nervous system arises from the neural tube, divisions of which will form the major structures with the hollow interior forming the spinal cord and ventricular system. This tube is lined by a single layer of neuroepithelial cells which will eventually give rise to all 3 neural lineages: neurons, astrocytes and oligodendrocytes. This occurs in a timely manner, with neurogenesis preceding astrogenesis, which precedes generation of oligodendrocytes. These processes occur during embryogenesis and early postnatal stages with only niche areas of neurogenesis remaining in the adult. At around E11 in the mouse the transcriptional profile of the neuroepithelial cells, located in the ventricular and subventricular zones, change and they transform to radial glial cells. These are elongated cells with apical and basal processes spanning the width of the cerebral wall from the ventricular cavity to the pial

surface. This allows them not only to act as a scaffold for migrating neurons, but also to receive important differentiation signals from the cerebrospinal fluid. They undergo asymmetric division to produce another radial glial cell and an immature neuron or intermediate progenitor cell which are capable of 1-3 self-renewal symmetric divisions before division to form 2 neurons. The mouse cerebral neocortex is made up of 6 horizontal layers. The formation of this neocortex occurs in an inside out manner, in which neurons destined for deep layers are generated before younger neurons which migrate past them to more superficial layers. In late embryogenesis and early postnatal stages the radial glial cells detach from the ventricular surface and transform into astrocytes and oligodendrocytes with some remaining to form the adult stem cell pool (Reviewed in(Okano and Temple, 2009; Yao et al., 2016)) (Figure 1.7).

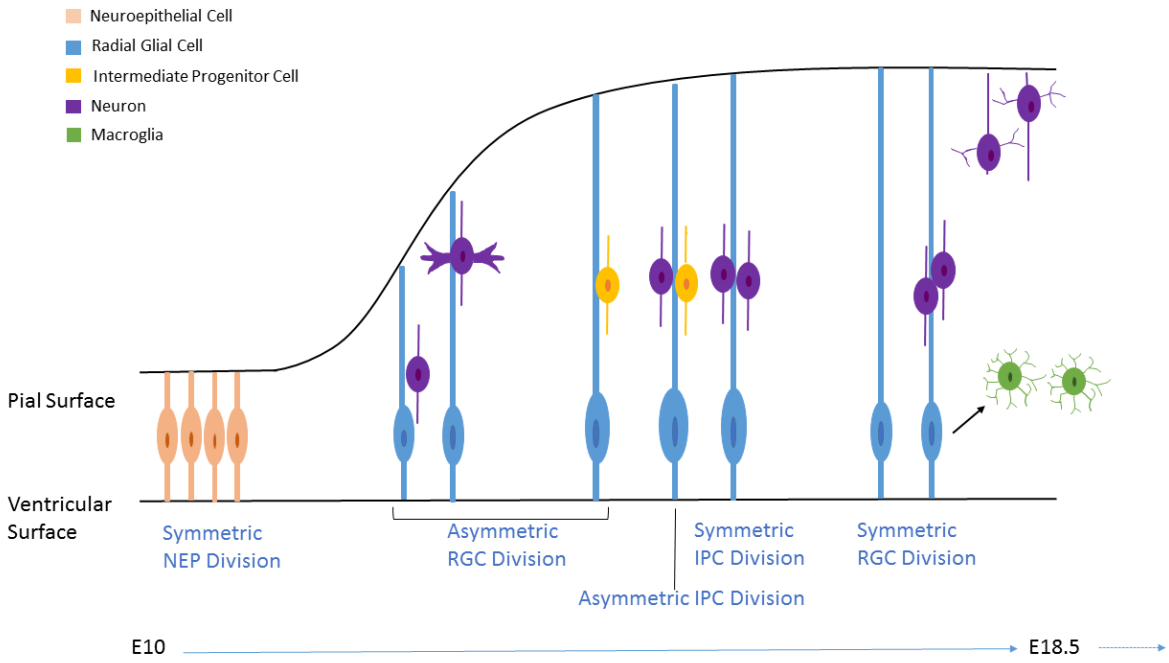


Figure 1.7. Embryonic Neurogenesis. In mice, neuroepithelial (NEP) cells lining the neural tube transform to radial glial cells (RGCs) around E11. These RGCs undergo asymmetric division to self-renew and form either a neuron or an intermediate progenitor cell (IPC). These newly formed cells use RGC processes as scaffolding to migrate to the appropriate cortical layer. IPCs have the capacity to undergo a limited number of asymmetric divisions before the generation of 2 neurons. Later in development RGCs form astrocytes and oligodendrocytes. Concept for figure from (Paridaen and Huttner, 2014).

If the steps of neurogenesis do not occur in a timely manner, abnormal brain size, cellular composition or malformations can result (Bansod et al., 2017; Desikan and Barkovich, 2016; Dwyer et al., 2016). Timing is therefore tightly controlled by a range of factors which include extracellular signalling pathways, intrinsic TF networks and epigenetic modifications. The importance of cell intrinsic programs is highlighted in cells differentiated in culture which

recapitulate the timing of neurogenesis and astrogenesis, making this a valid system for investigation of this process (Qian et al., 2000; Ravin et al., 2008; Shen et al., 2006).

1.6.3.2 Dynamic DNA methylation during neurodevelopment

One of the ways in which this timing of neurogenesis is controlled is through DNA methylation. The *Dnmts* are differentially expressed depending on the stage of neurogenesis. *Dnmt1* is ubiquitously expressed in NPCs and postmitotic neurons. *Dnmt3b* is expressed in early neural progenitor cells but is down-regulated after E13.5 being undetectable after E15.5. In contrast Dnmt3a is not detectable until later in development around E10.5-E13.5 and continues to be expressed in adult neurons (Feng et al., 2005; Watanabe et al., 2006).

Dnmt3a$^{-/-}$ NPCs in culture demonstrate down-regulation of neurogenic genes and up-regulation of genes involved in astrocyte and oligodendrocyte differentiation; these genes having lost their promoter methylation, resulting in 10-fold fewer neurons being produced (Wu et al., 2010; Wu et al., 2012). This premature glial differentiation is seen in other hypomethylated mutants. In *Dnmt1* KO NPCs methylation is lost from astroglial genes and genes essential in JAK-STAT signalling, a downstream target of which is activation of GFAP which is required for astrocyte differentiation, resulting in premature astrogenesis (Fan et al., 2005a).

Whilst cellular demethylation, remethylation programmes (CDRs) were previously thought only to occur in totipotent and germ cells, there is now evidence for this process occurring in post-mitotic neurons. Sharma *et al* examined methylation at different stages of neurogenesis in 2 cell lineages. Firstly in ES cells, secondly in dorsal and ventral telencephalic NPCs, then in immature and finally mature hippocampal and striatal neurons. They found an initial genome wide methylation during progenitor proliferation, which was followed by a loss of methylation upon the transition to immature neurons, and finally there was a further wave of methylation during maturation of these neurons. The wave of demethylation on progenitor to immature neuron transition corresponded to a peak in *Tet3* expression. Methylation and transcription were very similar between the 2 lineages at progenitor and immature neuron stages. Subtype-specific changes in methylation and transcription occurred during the maturation phase with methylation changes occurring preferentially at intergenic sequences with enhancer like activity at both CpG and non-CpG sites. This was blocked by conditionally inactivating *Dnmt3a* (Sharma et al., 2016).

The timing and extent of this CDR appears individual to each neuronal cell type. Another group reported that the most extensive CDR occurs in cerebellar Purkinje cells, which undergo an extensive CDR at P14-30 in the mouse brain, coinciding with dendritogenesis

and synaptogenesis (Zhou et al., 2016). As one of the largest neurons with the most extensive synaptic network, one might hypothesise that a wider scale reprogramming event would be necessary for their generation. Much smaller scale demethylation, remethylation events may also occur in response to neuronal activity (Guo et al., 2011; Martinowich et al., 2003).

1.6.4 Neurological disorders associated with abnormal DNA methylation

The importance of furthering our understanding of DNA methylation mechanisms during neurodevelopment is highlighted by the fact that several neurodevelopmental disorders, other than ICF syndrome discussed previously, are associated with abnormal deposition of DNA methylation during development or by mutations in proteins forming part of the methylation machinery. Examples of these are discussed below.

Rett syndrome is caused by mutations in the methyl binding protein MECP2. It is characterized by onset of severe intellectual impairment, loss of speech and motor skills, deceleration of head growth and repetitive stereotypies between 6-18 months of age (Neul et al., 2010). Although classed as a neurodevelopmental disorder, the phenotype can be reversed upon restoration of MECP2 (Guy et al., 2007) and removal of MECP2 in adult mice results in the same phenotype (McGraw et al., 2011; Nguyen et al., 2012), which suggests it is actually required for neuronal function throughout life. As mentioned earlier MECP2 is considered a repressor of transcriptional activity. Retrotransposon transcripts are moderately increased in MECP2 knockouts and increased L1 retrotransposition has been reported in neurological models of Rett syndrome (Muotri et al., 2010).

A group of neurodegenerative conditions are caused by mutations in *Dnmt1*, namely Hereditary and Sensory Autonomic Neuropathy (HSAN) and Autosomal Dominant Cerebellar Ataxia, Deafness and Narcolepsy (ADCA-DN) (Klein et al., 2011; Winkelmann et al., 2012). Although classified as separate entities, it is more likely they represent a spectrum of disease. Patients present in early adult life with a range of symptoms which can include neuropathies, hearing loss, cerebellar ataxias, psychosis and early onset dementia (Baets et al., 2015). Global hypomethylation has been found in peripheral blood DNA from these patients, although at modest levels, and pathway analysis revealed that genes with differentially methylated regions were associated with diverse neurological disorders (Sun et al., 2014). However DMRs only had between 5-10% reductions in methylation and, if hypomethylation was the main pathogenic mechanism, it does beg the question as to why the disease onset is in adult life. It could be argued that a relatively small reduction in methylation has no effect during development but has an accumulating impact on the nervous system over time. A second pathogenic mechanism may be due to misfolding and

aggregation of the mutant DNMT1 protein causing cellular stress and eventually death (Baets et al., 2015). However it could also be due to an unknown role for *Dnmt1* in postmitotic neurons.

Abnormal hypo/hypermethylation at certain sites has been associated with a variety of neurological disorders. For example Fragile X syndrome, an inherited cause of intellectual disability and autism, is caused by a trinucleotide repeat expansion in the *Fmr1* gene which encodes FMRP, a protein which plays several roles in the brain: including nucleocytoplasmic shuttling of mRNA, synaptic protein synthesis and dendritic localization of mRNA. These CGG repeat expansions become abnormally hypermethylated, resulting in silencing of the *Fmr1* gene (Bell et al., 1991). There is also preliminary evidence for aberrant methylation patterns in patients with psychosis (Xiao et al., 2014).

1.6.5 Mouse Models

In order to gain a better understanding of how abnormal DNA methylation during development can result in neurological disease, several mouse models have been generated. As previously mentioned full *Dnmt* and *Lsh* knockout models die prematurely. This means that one cannot study the consequences on the mature brain.

Models have therefore been generated with targeted knockout in the neural system. Fan *et al* generated mice in which *Dnmt1* was knocked out in NPCs using a *Nestin-Cre* model, or in postmitotic neurons using *CamK-Cre*. In NPCs this resulted in significant hypomethylation, including in postmitotic progeny. Mice in which there was a 95% recombination efficiency died within an hour following C-section delivery at E18.5-E19.5 due to respiratory failure, whilst those who had only a 30% recombination efficiency survived into adulthood. In these survivors the hypomethylated neurons were selected against and so eliminated from the brain during postnatal development. In contrast, knockout in postmitotic neurons did not affect methylation levels and had no effect on survival of mutant mice (Fan et al., 2001). Then again *CamK-Cre* drivers are selective for glutamatergic neurons, and in adult brain *Dnmt1* is mainly expressed in GABAergic neurons. Therefore one cannot exclude a role for *Dnmt1* mediated methylation in these postmitotic cells.

The same group then used the *Emx1-Cre* system to knock out *Dnmt1* in precursors of forebrain excitatory neurons. Resultant mice survived into adulthood but showed abnormal neurodevelopment with indistinct cortical lamination (Golshani et al., 2005). The hypomethylated neurons were selected against during development with severe neuronal death between E14.5 and 3 weeks postnatally causing cortical and hippocampal degeneration. Some hypomethylated neurons escaped this selection process accounting for 20-30% of neurons in the adult brain. Mutant neurons showed abnormal dendritic arborisation and excitability. Adult mice exhibited hyperactivity, learning and memory

defects. There was differential expression of 1500 genes including those involved in cortical layer specification, neural activity and cell death (Hutnick et al., 2009). The difficulty in interpreting this data is that as *Dnmt1* was not knocked out in inhibitory interneurons in this model, it is presumed there will be a greater percentage of these in mutant brains. This would have consequences upon transcriptome analysis and behaviour.

A *Dnmt3a*$^{-/-}$ *Nestin-Cre* mouse model showed evidence of hypo activity by 8 weeks and behavioural testing at 12 weeks demonstrated neuromuscular defects with abnormal gait, weakened grip strength and motor coordination and died prematurely at 18 weeks. Histological analysis suggested this was due to decreased numbers of motor neurons and fragmentation of endplates at neuromuscular junctions. IAP methylation, which was used as a proxy for global methylation, however appeared to be unperturbed (Nguyen et al., 2007).

Feng *et al* generated mice with targeted deletion of *Dnmt3a, Dnmt1* or both in postmitotic neurons using *CamK-Cre*. They found only double mutants demonstrated a phenotype. They had normal lifespan and did not display any obvious phenotype on observation in the cage. Histological examination revealed smaller hippocampal volume and abnormal synaptic plasticity. This correlated to deficits in learning and memory. Global methylation analysis revealed a 20% reduction in methylation in double mutant brains. A small number of genes (91) were misexpressed in the double mutant brains important in synaptic function, learning and memory. This suggests that *Dnmt1* and *Dnmt3a* may play redundant roles in postmitotic neurons (Feng et al., 2010a).

Targeted deletion of *Uhrf1* in an *Emx1-Cre* system whilst having little effect on embryonic development, resulted in global DNA hypomethylation with activation of IAPs and severe postnatal neurodegeneration. This study also revealed that perturbed DNA methylation caused by a factor only expressed in earlier stages of differentiation, can have long-reaching consequences beyond the time-point when the protein is expressed (Ramesh et al., 2016).

These models, whilst informative, are not ideal for studying the effects of hypomethylation during neurodevelopment on the adult brain due to premature death, lack of effect on DNA methylation or only a subset of neurons being targeted. A model in which methylation is perturbed from NPC stage that survives into adulthood, retaining hypomethylated neurons, would therefore be desirable in order to investigate molecular and phenotypic consequences of this and begin to unravel the aetiology behind associated neurological disorders.

1.7 Thesis aims

The wide scale remodelling of DNA methylation during embryogenesis to restore pluripotency, and then increasingly restrict cell fate during differentiation, is one of the most impressive examples of epigenetics in action. This is particularly true in the developing brain. Indeed, the disruption of this process is associated with various neurological disorders. Despite this, the mechanisms behind DNA methylation during this period are still to be fully understood. Furthermore, how disruptions in this process may lead to disease in the mature brain remains unknown largely as a result of the lack of available models.

Lsh, a chromatin remodeller, is known to be essential for DNA methylation during development. Mutations in this gene are also causative for ICF syndrome, a symptom of which is intellectual disability. Despite the fact that *Lsh* is down-regulated upon differentiation, the majority of research into *Lsh* function has been carried out in somatic cells. Furthermore, how mutations in this gene lead to neurological symptoms is unknown as knockout models die prematurely with multi-system failures, preventing analysis of the mature brain.

In this thesis I intend to utilise a neurodifferentiation cell culture system with $Lsh^{-/-}$ ES cells and a neurally targeted $Lsh^{-/-}$ mouse model to further our understanding of these unknowns. The main aims being as follows:

- To utilise a neurodifferentiation protocol to analyse the requirement for *Lsh* for normal neurogenesis and its contribution to DNA methylation during early neurogenesis.
- To analyse the methylome of $Lsh^{-/-}$ neural progenitor cells, taking a genome-wide approach as the majority of previous studies have focused only on selected sites.
- To analyse the transcriptome of *Lsh* deficient neural progenitors and its relation to DNA methylation defects.
- To characterise a novel mouse model in which *Lsh* knockout is targeted to neural tissue. This was done with the intention of generating a model in which the consequences of defects in DNA methylation during development upon the mature brain could be analysed.
- This characterisation will include genome-wide analysis of the methylome and transcriptome in the $Lsh^{-/-}$ mouse brain. Something which has not been reported on before. It is hoped this may reveal insights into potential pathology underlying the neurological defects seen in ICF syndrome.

Chapter 2. Materials and Methods

2.1 Cell culture

In order to model embryonic development, one has to start with a culture of pluripotent cells. These cells have the ability to differentiate into any of the three primary germ layers, and so, can contribute to any tissue in the developing organism.

The most commonly used pluripotent stem cell model is that of mouse ESCs derived from the ICM of the blastocyst. When maintained in appropriate media these cells have the ability to self-renew, making them ideal for propagation of long term stable cultures and for genetic manipulation to create transgenic lines (Evans and Kaufman, 1981; Martin, 1981). Appropriate media contains foetal calf serum, a key component of which is bone morphogenic protein 4 (BMP4). This signalling molecule results in expression of *Inhibition of differentiation* genes via SMAD signalling pathways thereby aiding to suppress differentiation in culture (Ying et al., 2003a). Another important component of mESC culture media is Leukaemia Inhibitory Factor (LIF). This cytokine stimulates expression of pluripotency genes by activating JAK/STAT pathways (Hirai et al., 2011). mESCs, even after extensive culture, have the ability to differentiate to cells of all three germ layers and when injected into the blastocyst can contribute to all tissues in the developing embryo (Bradley et al., 1984). I have therefore selected this system to model early neurogenesis.

2.1.1 mESC culture

All cells were grown in cell culture grade tissue culture flasks or plates (ThermoFisher Scientific) coated in 0.1% gelatin.

mESC media was composed of Glasgow Modification Eagles Medium (GMEM, Gibco) supplemented with 15% foetal calf serum (HGU technical services), 1% sodium pyruvate (Sigma), 1% non-essential amino acids (Sigma), 1000 U/ml penicillin and 650 µg/ml streptomycin (HGU technical services), 1% L-glutamine (HGU technical services), 0.1 mM β-mercaptoethanol (Invitrogen) and 500 units/ml of LIF (Millipore) and stored in a humidified atmosphere at 37^0C and 5% CO_2. Media was changed every 1-2 days.

Cells were passaged every 2-3 days when reaching 80% confluency. This was done by rinsing cells with PBS (Sigma) following removal of media to remove traces of serum. Cells were then incubated with 1x trypsin-EDTA v/v (Sigma) at 37^0C for 2 minutes. Cells were dislodged by gentle agitation and trypsin inactivated by the addition of serum containing media. The sample was then centrifuged for 5 minutes at 300 x g. After removal of media the cell pellet was then resuspended in fresh media and split into new flasks.

Supernumerary cells were rinsed in PBS, centrifuged and resuspended in CryoStor cell cryopreservation media (Sigma) in cryovials. These were then placed in a “Mr Frosty” (ThermoFisher Scientific) at 4^0C for 15 minutes, then placed at -80^0C to allow a controlled reduction in temperature of 1^0C per minute for 24 hours, for optimal cell preservation, before transfer to liquid nitrogen for long term storage.

On recovering cells from liquid nitrogen storage, they were rapidly thawed in a 37^0C water bath, then added drop wise to culture media, centrifuged for 5 minutes at 300 x g and the cell pellet resuspended in fresh media.

2.1.2 Generation of cell lines

E14 mESCs were used throughout this project as a wild type control cell line. *Lsh*$^{-/-}$ mESCs were generated by Dr Donncha Dunican using the CRISPR/Cas9 gene editing system. E14 mES cells were transfected with Cas9 and a single guide RNA designed to target exon 1 of the *Lsh* gene. XhoI restriction digest was used to screen for clones containing a mutation. The target site was amplified by PCR and subjected to Sanger sequencing to establish the nature of the mutation. The clone I have used was shown to have a homozygous 25bp deletion in exon 1 (Figure 2.1).

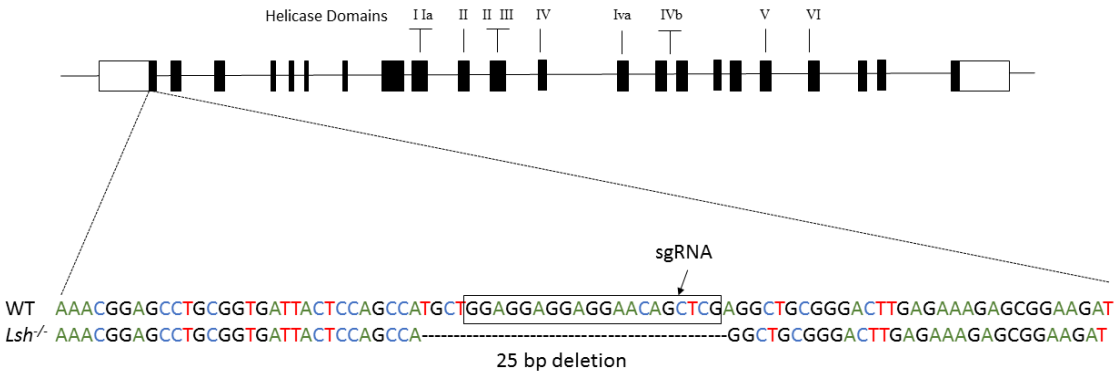


Figure 2.1. Schematic of CRISPR targeted deletion. The 25bp deletion in exon 1 present in CRISPR generated *Lsh*$^{-/-}$ cells is represented in the diagram along with the single-guide RNA target site.

From these *Lsh*$^{-/-}$ mESCs, I generated two rescue cell lines. One line named *Lsh*$^{-/-(+\ WT\ Lsh)}$ was rescued with a *PiggyBac* vector encoding full length wild type *Lsh.* Another line named *Lsh*$^{-/-(+\ Mutant\ Lsh)}$ was rescued with a *PiggyBac* vector encoding *Lsh* carrying a mutation in a conserved helicase domain (R699Q). The rescue cell lines were generated by transfecting *Lsh*$^{-/-}$ mESCs with *PiggyBac* vectors designed by Professor Meehan and produced by VectorBuilder.

For transfection, *Lsh*$^{-/-}$ mESCs were plated onto gelatinised 6 well plates at a density of $5x10^5$ per well in Penicillin/Streptomycin free serum media. 2.5 µg of vector DNA in 50 µl of

Opti-MEM (ThermoFisher Scientific) with 6 µl lipofectamine (Invitrogen) and 1 µg of transposase vector (Transposagen) was combined, vortexed and left at room temperature for 20 minutes. This was then added drop wise to each well. Media was changed after 5 hours and then again after a further hour using ES media containing Penicillin/Streptomycin with the addition of Puromycin 2 µg/ml in cells transfected with WT *Lsh* vector, or Hygromycin B 200 µg/ml in cells transfected with the mutant *Lsh* vector according to their selectable marker. Cells were grown under these conditions for 10 days and viable cells were then single cell sorted via FACS and clones expanded. Successful transfection was determined by the presence of LSH on western blotting.

2.1.3 Neurodifferentiation

The neurodifferentiation protocol used was adapted from Pollard *et al* (Pollard et al., 2006). This protocol is thought to effectively model neurogenesis in the embryo. Initially, cells start to form clusters termed “neural rosettes”. These rosettes are said to resemble the neural tube of the embryo, as the neuroepithelial cells forming these rosettes display the typical apico-basal polarity of the cells of the neural tube and active notch signalling, both of which are crucial to both maintain a population of proliferating neural progenitors, and to coordinate the timely production of neurons and glia. Mature neural cells can then be seen at the periphery of these rosettes and neurogenesis precedes gliogenesis (Abranches et al., 2009).

On the day prior to differentiation, ES cells were split at high density to reduce the numbers of differentiated cells carrying through. On the day of differentiation, ES cells were dissociated from flasks using Trypsin and rinsed twice in PBS to remove any traces of serum which would be inhibitory for differentiation. Cell pellets were then resuspended in neurodifferentiation media consisting of a 1:1 mix of Neurobasal (Gibco) and DMEM/F-12 media (Gibco) with 0.5% Neuro-2 Supplement (Millipore), 0.5% B27 Supplement (Gibco), 0.2 mM L-glutamine and 0.1 mM β-mercaptoethanol. They were then plated onto 0.1% gelatin-coated flasks (or glass slides for use in immunocytochemistry) at a density of $3x10^5$ cells per T75 flask. Media was changed on day 2 of differentiation and then daily thereafter.

2.1.4 Cell harvesting

Cells were harvested by dissociation with Trypsin and centrifuged at 300 x g for 5 minutes, washed in PBS and centrifuged again. The resulting cell pellet was either snap frozen on dry ice for later DNA or protein extraction, or resuspended in TRIzol (Invitrogen) and snap frozen for later RNA extraction. Samples were then stored at -80^0C.

When cell counting was required, such as prior to neurodifferentiation, cells were harvested and resuspended in media. Trypan blue 0.4% (ThermoFisher Scientific), which stains non-viable cells, was added to this suspension in a 1:1 ratio. 10 µl of this solution was then loaded onto a haemocytometer. The number of viable cells across all 4 quadrants was

counted and an average then calculated, multiplied by 2, to account for Trypan blue dilution, and multiplied by $1x10^4$ to give the number of cells per ml of suspension.

2.1.5 Magnetic Activated Cell Sorting

In order to select for neural progenitor cells in heterogeneous cultures, cells underwent Magnetic Activated Cell Sorting (MACS) after 10 days of culture in neurodifferentiation media. Anti-PSA-NCAM Microbead kit (Miltenyi Biotec) was used as per manufacturer's instructions. Polysialylated neural cell adhesion molecule (PSA-NCAM) is a marker for neural progenitor cells.

A single cell suspension was obtained, following trypsinisation, by passing cells through a 30 µm nylon mesh and resuspending in buffer (0.5% bovine serum albumin and 2 mM EDTA in PBS). Cells were then incubated with Anti-PSA-NCAM Microbeads 20 µl per $1x10^7$ cells for 15 minutes at 4°C. Following this, the cells were washed in 2 ml of buffer, centrifuged at 300 x g for 10 minutes and the pellet resuspended in 500 µl of buffer. This was then passed through an LS MACS Column within a "MidiMACS" magnetic separator. This column was washed three times with buffer to flush through PSA-NCAM negative cells. Cells retained in the column were then eluted by removing the column from the magnetic separator and flushing through with buffer. This process was repeated over a second column to increase the purity of eluted cells.

2.1.6 FACS Counting

In order to quantify neural progenitor cell generation and subsequent purification by MACS, I employed Fluorescence-activated cell sorting (FACS) using anti-PSA-NCAM antibody conjugated to PE (Miltenyi Biotec). As per manufacturer's instructions, $1x10^6$ cells were resuspended in 100 µl of buffer and 10 µl of antibody added. The suspension was then stored in the dark at 4°C for 10 minutes and then washed twice in 2 ml of buffer. The cell pellet was then resuspended in 200 µl PBS. This process was carried out from WT and *Lsh*$^{-/-}$ ESCs and day 10 differentiated cells pre and post MACS. This was done on 3 biological replicates for each condition.

The positive fraction in each sample was quantified on an Aria FACS machine with the aid of Elisabeth Freyer (HGU technical services). The positive population was determined according to negative controls of undifferentiated ESCs, and day 10 differentiated cells with no antibody. An explanation of the gating tree used is shown in Figure 2.2. A forward scatter height (FSC-H) vs. forward scatter area (FSC-A) density plot was used to determine gate P1 for selection of single cells. A forward scatter vs side scatter gate (P2) was then used to identify cells of interest and exclude debris. Gate P3 was used to determine the population positive for PE fluorescence. Data was then displayed on histograms to show the overall

count of positive and negative cells. Percentage of positive cells was determined as the percentage of cells in gate P2 that were in gate P3 (Figure 2.2).

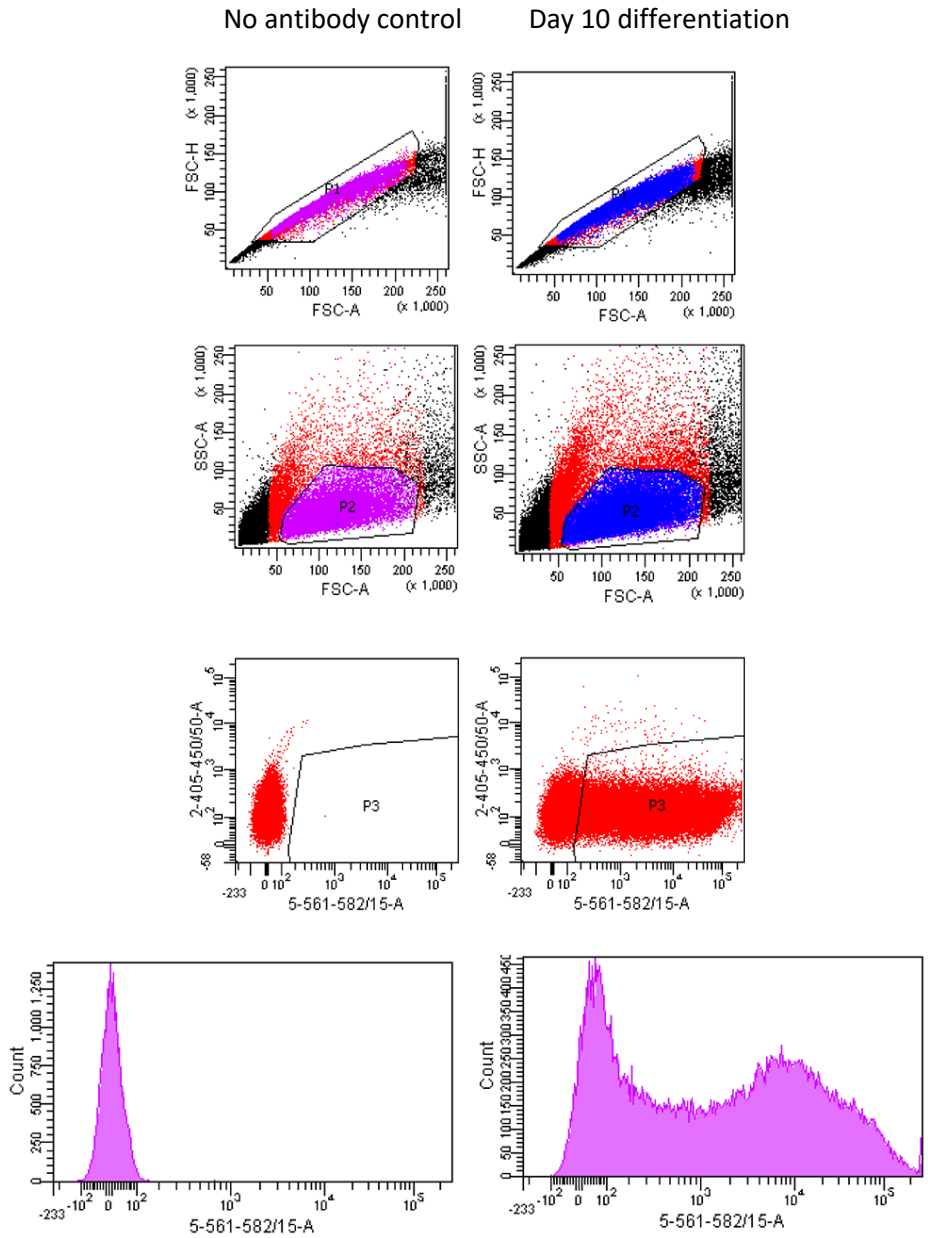


Figure 2.2. FACS analysis. Representative plots of FACS experiments to demonstrate gating used to analyse samples. Plots shown are from a WT culture following 10 days of neurodifferentiation stained with anti-PSA-NCAM-PE and a control sample of the same type with no antibody staining. Gate P1 was set by plotting forward scatter height by forward scatter area to select for single cells. Gate P2 was set by plotting forward scatter vs side scatter to select cells of appropriate size and remove debris from the analysis. Gate P3 selects the positive population based on PE fluorescence. Histograms demonstrate the positive and negative populations.

2.2 Cell Imaging

2.2.1 Phase contrast imaging

Phase contrast images of cell cultures were taken by phase contrast microscopy using the Nikon Eclipse Ti-S inverted microscope at 10x objective with iVision software.

2.2.2 Immunofluorescence

For immunocytochemistry experiments, cells were grown on gelatin coated glass coverslips. Media was removed and cells were washed in PBS and fixed in 4% paraformaldehyde (PFA) (Sigma) for 10 minutes at room temperature. PFA was then removed and fixed cells were washed twice in PBS and blocked in 10% Donkey serum (Sigma) in 0.1% Triton-X (Sigma) for 90 minutes. After blocking, cells were then incubated with primary antibodies (Table 2.1) diluted in 1% donkey serum in 0.1% Triton-X overnight at 4^{o}C. Cells were subsequently washed three times for 10 minutes in PBS, and then incubated with AlexaFluor-555 fluorophore-conjugated secondary antibodies (ThermoFisher Scientific) diluted in 1% donkey serum in 0.1% Triton-X at a concentration of 1:200, for 50 minutes in the dark. Cells were washed 3 times with PBS for 10 minutes in the dark and then stained with the nuclear counterstain DAPI (4'6-diamidino-2-phenylindole 50 µg/ml diluted 1:1000 in PBS) for 10 minutes in the dark. Coverslips were then washed in PBS twice for 10 minutes and once in distilled water. Coverslips were then transferred to Superfrost glass slides (ThermoFisher Scientific) and mounted with Vectashield mounting media (Vector laboratories). Image capture and analysis was performed using a Zeiss Axioscope 2 fluorescence microscope at 10x objective and Micro-manager software. Control slides were treated with secondary antibody only.

Primary antibody	Supplier	Concentration	AlexaFluor-555 secondary antibody
Nanog	ABCAM; Ab80892	1:75	donkey-anti-goat (A21432)
Nestin	ABCAM; Ab6142	1:200	donkey-anti-mouse (A31570)
Tuj	ABCAM; Ab14545	1:1000	donkey-anti-mouse (A31570)

Table 2.1. Antibodies used for immunofluorescence studies

2.2.3 Metaphase spreads

Metaphase spreads were prepared from cell cultures for assessment of karyotype as follows. Colcemid 1 µg/ml was added to cell media 30 minutes before harvesting. The cell pellet was then resuspended in 10 ml hypotonic solution (0.5% trisodium citrate, 0.25% potassium chloride in dH2O) and incubated at 37 ^{o}C for 10 minutes. The solution was then centrifuged

and the cell pellet resuspended in 3:1 methanol: acetic acid fix solution. The solution was centrifuged and pellet resuspended in fix solution a further 3 times before being stored at -20 ^{o}C. Solution was then dropped onto glass slide at arm's length and stained with DAPI 50 µg/ml diluted 1:1000 in PBS for 3 minutes. Slides were then imaged using a Zeiss Axioscope 2 fluorescence microscope at 100x objective.

2.3 Mouse model

2.3.1 Generation

The mouse model used in this project was generated, bred and genotyped by David Read and Dr Ian Adams.

A schematic representation of the generation is shown in Figure 2.3. Two-cell mouse embryo straws, heterozygous for a transgenic knock-out-first allele were purchased from The European Conditional Mouse Mutagenesis Programme, and implanted into a pseudo-pregnant mouse of C57BL/6 background. The transgenic allele consists of a gene trap cassette containing a splice acceptor site and a downstream transcriptional termination sequence. This creates a mutant allele encoding a truncated and non-functional protein. The gene trap cassette is flanked by flippase recognition target sites (FRT). The transgenic allele also contains Lox-P sites around exon 12 of the *Lsh* gene. Transgenic heterozygous offspring were then crossed with mice expressing flp recombinase which excises the gene trap cassette at FRT sites essentially creating a functionally WT allele that is floxed around exon 12 which encodes the conserved helicase domain IV. Heterozygous offspring from this cross are then crossed with mice expressing *Cre* driven by the *Nestin* promoter. This creates heterozygous offspring carrying the floxed transgene and expressing *Cre* in neural progenitor cells, which can then be further crossed to result in offspring homozygous for the mutant allele in neural cells. PCR genotyping was carried out by David Read using DNA derived from ear clips.

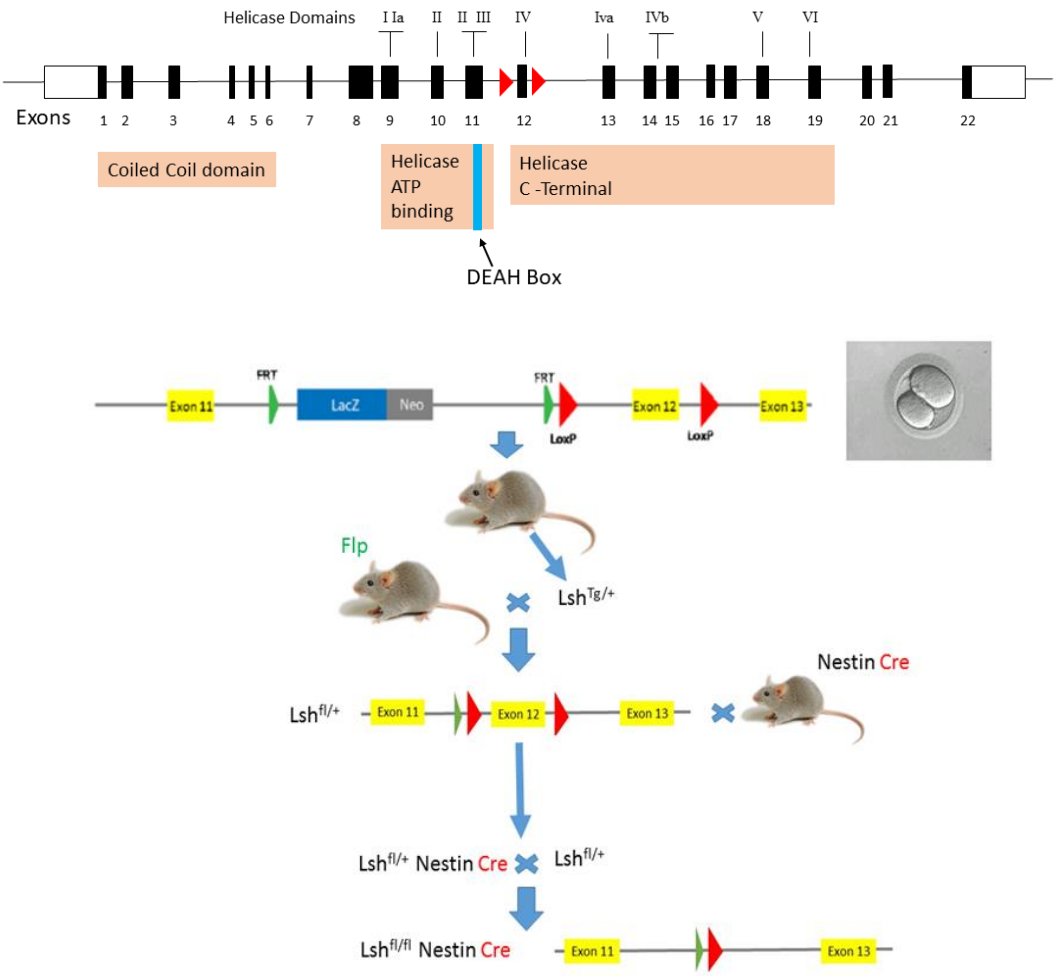


Figure 2.3. Generation of the mouse model. Schematic representation depicting generation of the targeted *Lsh* knockout mouse model. See text for detailed description. Green triangles represent flippase recognition target sites (FRTs) and red triangles LoxP sites. ^{Tg} depicts the transgenic allele, ^{Fl} the transgenic allele following flp recombination and $^{+}$ the wild type allele.

Animal care was in accordance with institutional guidelines and UK Home Office regulations. Mice were housed in environmentally enriched cages with littermates and maintained on a 12 hour light/dark cycle with ad libitum food and water. Experiments were carried out in accordance with the European Communities Council Directive (86/609/EEC) and a project licence with approval under the UK Animals Scientific Procedures Act (1986).

2.3.2 Sacrifice and dissection

Mice were sacrificed at 12 weeks or 1 year of age. Sacrifice was carried out using CO_2 gas administration and cervical dislocation was performed secondarily to ensure euthanasia. Dissection was then performed by David Read with removal of the brain, kidneys, spleen and liver. Cerebellum was separated from the forebrain, and left and right hemispheres of both forebrain and cerebellum were dissected. Left hemispheres, and other organs to be used for later nucleic acid extraction, were immediately snap frozen in liquid nitrogen and

then stored at -80 °C. Right hemispheres, for use in imaging, were fixed in 4% PFA at 4 °C for 48 hours. They were then placed in 20% sucrose solution for dehydration at 4 °C until they sank, before being embedded in Optimal Cutting Temperature Compound (OCT, Fisher Scientific) and stored at -80 °C. Brain weights were obtained using analytical balance APX60 (Denver Instruments) and corrected for total body weight.

For protein and nucleic acid extraction, snap frozen tissues were crushed in a mortar and pestle over ice and then pipetted in 4 °C PBS to dissociate tissue to single cells. Samples were then centrifuged and PBS removed to achieve a cell pellet.

2.3.3 Imaging

Sagittal tissue sections were cut from OCT embedded brain using a Leica Cryostat at 20µm thickness and transferred to subbed glass slides. Subbed slides were prepared by immersing Superfrost slides in 0.5% gelatin for 2 minutes and then drying at 50 °C overnight. Sections were then stored at -80 °C.

Cresyl violet staining was used to assess morphology. This stains Nissl substance (the rough endoplasmic reticulum) in the cytoplasm of neurons purple/blue due to ribosomal RNA staining, and is one of the most commonly used stains to assess brain structure. Cresyl violet stock solution was prepared by adding 0.2 g of cresyl violet acetate (Sigma) to 150 ml dH20 and stirring for 2 hours at room temperature. Stock solution was then diluted 1:10 with buffer solution (282 ml 0.1M acetic acid, 18ml 0.1M Sodium Acetate) and mixed at room temperature for 1 hour. Brain sections were thawed to room temperature and then immersed in Xylene for 5 minutes, 95% ethanol for 3 minutes, 70% ethanol for 3 minutes, dH20 for 3 minutes, cresyl violet solution for 8 minutes at 60 °C, dH2O for 3 minutes, 70% ethanol for 3 minutes, 95% ethanol for 1 minute, dipped once in 100% ethanol and finally immersed in xylene for 30 minutes. They were then mounted using xylene based mounting media and imaged using the Olympus Dotslide microscope at 20x magnification.

Images were analysed using Dotslide VS-ASW software. Measurements were taken from tissue sections from 12 week old male mice aligning to plate 14 (forebrain) and plate 9 (cerebellum) of the Allen Mouse Brain Atlas (mouse.brain-map.org). Sections from 3 mice in each genotype were studied. For each mouse 3 sections were analysed and each measurement on each section was repeated 3 times. Therefore for each mouse, each measurement was calculated from an average of 9 measurements in total. Measurements taken are illustrated on Figure 2.4. Measurements of cortical thickness were taken from the pial surface to the edge of the corpus callosum. For somatosensory cortex the anterior point of the hippocampal formation was used as a reference, and for the somatomotor cortex a point just before the curve of the genu of the corpus callosum was used as a reference point.

The area of the hippocampal formation and of the granular layer of the dentate gyrus was also taken.

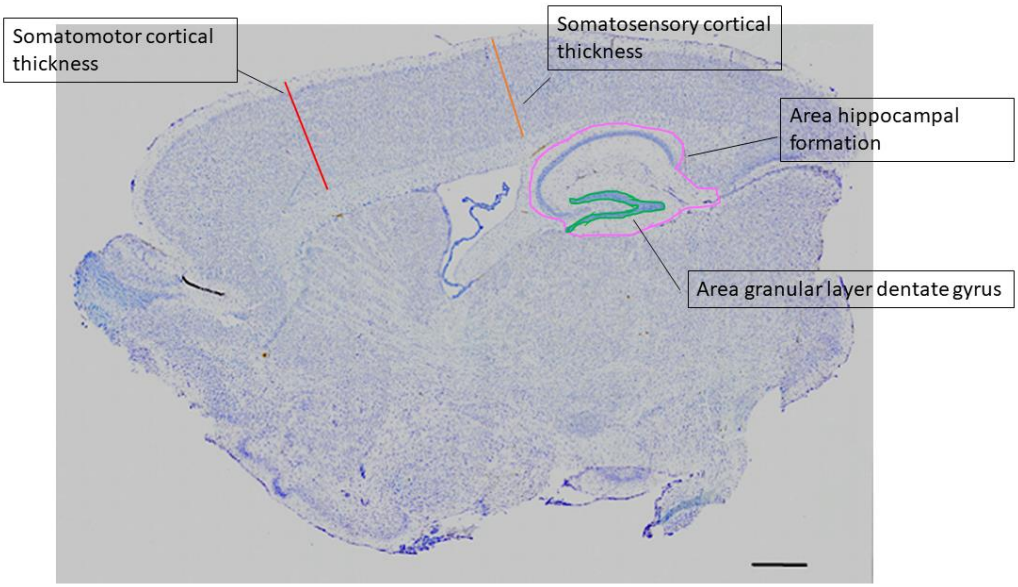


Figure 2.4. Mouse brain measurements. A representative sagittal section of a 12 week old mouse brain with the sites of measurements highlighted. Scale bar=500µm.

2.3.4 Behavioural studies

Behavioural studies were carried out in 1 year old male mice. 3 knockout subjects and 3 littermate controls were used. Experiments were carried out between 10 am and 2 pm. Mice were handled for 2 days prior to experimentation. Experiments were carried out in a quiet room with the light level set at 50 lux. Mice were transferred to the experimental room 1 hour prior to testing to allow acclimatisation. Experiments were recorded for later analysis. Equipment was cleaned with disinfectant between each mouse.

2.3.4.1 Open field

Open field test was carried out in a 50x50x38cm opaque Perspex box with a 16 12.5x12.5 cm grid on the base. Mice were gently placed in the central square and allowed to freely explore the box. Behaviour was recorded for 15 minutes. The test was repeated on the following day as this allows one to assess novel environment exploration on the first test and habituation on the second. Ambulatory measurements were made by measuring entrance into a new square on the grid. A mouse was considered to have entered a new square when all 4 paws had crossed a boundary point.

2.3.4.2 Y maze

Spontaneous alternation test was carried out in a Perspex Y maze with 35cm long arms, 5cm in width and 10cm high. The mouse was gently placed in the centre of the maze and allowed to move through the maze freely for 10 minutes. Mice with intact spatial and working memory show a preference to explore a less recently visited arm. This can be quantified by calculating the percentage of spontaneous alternation which is the number of successful triads (i.e. each of the 3 arms entered in turn) divided by the total number of possible triads (Miedel et al., 2017). A mouse was considered to have entered an arm of the maze if all 4 paws passed the boundary point.

2.4 DNA extraction and analysis

2.4.1 DNA extraction from cells and tissue

Cell pellets from tissue and cell lines were resuspended in 1 ml of DNA lysis buffer (10mM Tris pH8.0; 100mM NaCl; 10mM EDTA pH8.0; 0.5% SDS) and 5 µl of 20 mg/ml proteinase K (ThermoFisher Scientific). Samples were placed in a thermo-shaker at 55 ^{o}C, 1000 rpm overnight. Following this, samples were brought to room temperature and 2 µl of RNase cocktail (Ambion) added. They were then incubated in a thermo-shaker 37 ^{o}C, 1000 rpm for 30 minutes. An equal volume of Ultrapure Phenol:Chloroform:Isoamyl Alcohol (PCI, 25:24:1 v/v, Invitrogen) was then added to each sample which was vigorously shaken and centrifuged at 16,000 xg at 4 ^{o}C for 10 minutes to separate the aqueous phase containing the DNA from the organic phase containing protein and cell debris. The aqueous phase was then transferred to a new vial and the PCI extraction process repeated. An equal volume of chloroform was then added to the aqueous phase to remove phenol residues, vigorously shaken and centrifuged at 16,000 x g at 4 ^{o}C for 10 minutes before the aqueous phase was transferred to a fresh vial. To this vial 2.5x volume of 100% ice-cold ethanol and 1/10 volume of 3M sodium acetate was then added, and the sample was placed at -20 ^{o}C for a minimum of 2 hours to allow the DNA to precipitate. Samples were then centrifuged at 16,000 x g at 4 ^{o}C for 5 minutes to aggregate a DNA pellet. Supernatant was removed and the pellet carefully washed in 70% ethanol before being centrifuged again at 16,000 x g at 4 ^{o}C for 5 minutes. The supernatant was removed and the DNA pellet air dried before resuspension in TE buffer (100mM Tris pH8.0, 10mM EDTA pH8.0), or mass spectrometry grade water (Sigma) and stored at -20 ^{o}C. The concentration and purity of nucleic acids were measured using a Nanodrop ND-1000 spectrophotometer.

2.4.2 Gel electrophoresis

UltraPure agarose (Invitrogen) gels were made using 1xTBE (Tris/Borate/EDTA) or 1xTAE (Tris/Acetic acid/EDTA) with the addition of 1 µg/ml ethidium bromide to resolve nucleic acids. DNA samples were loaded using 1x Orange G loading buffer (50% glycerol v/v, 5mM

EDTA pH8.0, 0.3% orange G w/v). RNA samples were loaded with 2x RNA loading dye (ThermoFisher Scientific) following a 2 minute incubation at 80 ^{o}C. The size of the nucleic acid fragments was determined using GeneRuler DNA ladder mix (ThermoFisher Scientific) for DNA or RiboRuler (ThermoFisher Scientific) for RNA. Gels were visualized using a UV trans-illuminator or a FLA-5100 phosphoimager (Fuji).

2.4.3 Restriction digests

For digestion by restriction enzymes, 750 ng of DNA was added to a solution containing 10 units of restriction enzyme (either methylation sensitive HpyCH4IV (NEB) or methylation insensitive MSPI (NEB)) and 1.5 µl of CutSmart buffer (NEB) in a 15 µl reaction. This reaction was incubated at 37 ^{o}C for 90 minutes. The digested DNA was then resolved on a 1.5% agarose-TBE gel, run at 100 volts for 1.5 hours, and post stained in a 1 µg/ml ethidium bromide solution and rinsed in dH2O. Imaging of the gel was carried out on a FLA-5100 phosphoimager (Fuji).

2.4.4 Liquid chromatography mass spectrometry

Liquid chromatography mass spectrometry (LC-MS) was used to determine global 5mC levels. LC-MS is considered the gold-standard technique for quantifying global DNA methylation levels (Chowdhury et al., 2017). In this method, modified and unmodified cytosine bases can be separated according to their size and charge and then quantified.

1 µg of DNA was hydrolysed to single nucleotides with 10 units DNA Degradase Plus (Zymo Research) and 2.5 µl DNA degradase buffer (Zymo Research) in a 25 µl reaction with mass spectrometry grade water. Samples were incubated at 37 ^{o}C for 2 hours, then the enzyme was heat inactivated by incubation at 70 ^{o}C for 20 minutes. Samples were then centrifuged at 16,000 xg for 15 minutes. The upper 10 µl of solution was transferred to a fresh tube leaving behind the enzyme spun down to the bottom of the vial. To this fresh tube was added 100 µl mass spectrometry grade methanol and 60 µl mass spectrometry grade acetonitrile (Sigma). Samples were then centrifuged at 16,000 x g for 5 minutes, the upper 30 µl of supernatant transferred to mass spectrometry glass vials and submitted to the IGMM mass spectrometry facility. Standards for unmodified cytosine, 5mC and 5hmC were ran alongside the samples for comparison. LC-MS was performed by Jimi Wills.

For liquid chromatography, samples were separated on a SeQuant ZIC-pHILIC column using a Thermo UltiMate 3000 BioRS with a flow rate of 0.3ml/min and a gradient of 90% to 5% acetonitrile in 10 minutes. Eluting peptides were analysed on a Thermo Q Exactive in negative mode, scanning from 300 to 350 m/z at resolution 70k. AGC target was set to 3 x 10^{6} and maximum ion injection time 500 ms.

Data was analysed using the Xcaliber programme. Levels of 5mC were taken as a percentage of total cytosines (i.e. unmodified, methylated and hydroxymethylated).

2.4.5 Bisulphite sequencing

Bisulphite sequencing was employed to quantify DNA methylation at particular sites. Bisulphite treatment of DNA converts unmodified cytosines to uracil whilst not affecting methylated cytosines. Following PCR amplification and sequencing, converted uracils will be recognised as thymines allowing their distinction from modified cytosines which will remain as cytosine on sequencing. I used this method to quantify DNA methylation at major satellite repeats in ESCs and NPCs.

500 ng of DNA was bisulphite treated using EZ DNA Methylation-Lightening kit (Zymo Research) according to the manufacturer's instructions. Each sample was made up to 20 µl with dH2O and 130 µl lightening conversion reagent added. Samples were then incubated at 98 ^{o}C for 8 minutes then 54 ^{o}C for 1 hour. 600 µl M-binding buffer was added to a spin column and the sample was then loaded and mixed by inversion. Columns were then centrifuged at 16,000 x g for 30 seconds. Flow through was discarded and 100 µl M-wash buffer was added to the column. Columns were again centrifuged at 16,000 x g for 30 seconds and 200 µl of L-desulphonation buffer was added to the column and allowed to stand at room temperature for 15 minutes. Samples were spun for 30 seconds, 200 µl of wash buffer was added and centrifugation repeated. This wash step was repeated. The column was then placed onto a collection tube, 10 µl of M-elution buffer added and the sample centrifuged. The eluted DNA was stored at -80 ^{o}C.

PCR amplification of the major satellite region of interest was carried out as below. Major satellite primers were taken from Dunican *et al* (Dunican et al., 2013).

PCR reaction:

- 1.5 µl bisulphite treated DNA
- 2.5 µl forward and reverse primer mix (10 µM of each primer, see Table 2.2 for sequence)
- 5 µl 10x PCR buffer (Invitrogen)
- 5 µl 2mM dNTP mix
- 1.5 µl Mg^{2+} (Invitrogen)
- 0.5 µl Taq polymerase (Invitrogen)
- 34 µl dH20

Reactions were placed in a thermal cycler at 94 ^{o}C for 3 minutes followed by 39 cycles at 94 ^{o}C for 45 seconds, 57 ^{o}C for 30 seconds and finally 72 ^{o}C for 30 seconds.

DNA was precipitated by the addition of 5 µl 3M sodium acetate, 1 µl glycogen and 138 µl ice cold 100% ethanol and stored at -20 ^{o}C for 30 minutes. Following centrifugation at 16,000

x g for 10 minutes at 4 °C, the DNA pellet was washed in 70% ethanol, centrifuged and air dried before being resuspended in 6 µl dH2O.

The PCR products were then resolved on a 1.5% TAE gel with ethidium bromide, ran at 100 V for 1 hour. Bands were visualised using a Safe Imager Trans illuminator (Invitrogen). PCR products of the expected size (360bp) were cut from the gel using a scalpel. DNA was extracted from the agarose using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions: Gel sample weight was measured and 300 µl of buffer QG per 100 mg of gel was added. The sample was then incubated at 50 °C for 10 minutes in a thermo-shaker until the gel dissolved. 100 µl of isopropanol per 100 mg of gel was then added and the solution placed in a quick spin column and centrifuged at 16,000 x g for 1 minute. Flow through was discarded and 500 µl buffer QG added to the column and centrifuged again. 750 µl of wash buffer PE was then added and allowed to stand for 5 minutes before centrifugation. The spin column was then placed in a collection tube and 50 µl of buffer EB added to column. This was allowed to stand at room temperature for 1 minute before centrifugation and collection of the eluted DNA.

PCR products were cloned into the pGEM-T Easy vector (Promega). This is a linearized vector with a thymidine overhangs. PCR products generated with Taq polymerase leave adenosine overhangs thereby improving the efficiency of the ligation as the DNA has something to "stick" to.

Ligation reactions were set up with an insert to vector ratio of 3:1. Appropriate volume of insert was added to the reaction containing 5 µl 2x rapid ligation buffer, 1 µl PGEM T Easy, 2 µl DNA ligase, and made up to a 10 µl reaction with dH20. This was incubated at 4 °C overnight. Control experiments were set up using the supplied Control Insert DNA as a positive control and as a negative control no DNA was added to the ligation reaction.

Ligations were then transformed into library efficient DH5α competent *Escherichia Coli* cells (Invitrogen). 1 µl of ligated DNA was added to 40 µl of competent cells and incubated on ice for 30 minutes. Samples were then heat shocked in a 42 °C water bath for 45 seconds and immediately placed back on ice for 5 minutes. 500 µl of SOC media was added to each transformation and placed in a shaking incubator at 37 °C, 200 rpm for 1 hour. One third of each mixture was then aseptically spread onto L-AMP plates (L-agar plates containing 50 µg/ml ampicillin) which had been spread with 80 µl of 20 mg/ml X-gal (Sigma) and allowed to air dry prior to use. Inverted plates were then incubated at 37 °C to allow bacterial colony growth.

The pGEM-T Easy vector also encodes β-galactosidase. Successful cloning of an insert into the vector disrupts the sequence of this gene. β-galactosidase cleaves X-gal to form 5-bromo-4-chloro-indoxyl, which then oxidizes to form 5,5'-dibromo-4,4'-dichloro-indigo, which

appears bright blue in colour. This allows the identification of colonies containing an insert which appear white, whilst colonies lacking an insert will appear blue. The white colonies were then picked and transferred to a 96-well mini-prep plate, each well containing 1ml Luria-Bertani culture broth with 50 µg/ml ampicillin. Plates were placed in a shaking incubator set at 200 rpm, 37 ^{o}C overnight to allow colony expansion.

HGU technical services then isolated plasmid DNA and carried out Sanger sequencing. pGEM-T Easy vectors also contain T7 and Sp6 RNA polymerase promoters flanking the cloning region therefore Sanger sequencing was performed using T7 and Sp6 primers. This was done using the BigDye Terminator V3.1 sequencing kit, and samples were run on an ABI Prism 3720 genetic analyser. Sequences were then aligned to a reference sequence for the major satellite region and analysed using the BioEdit Sequence Alignment Editor V7.2.5. Methylation status of each CpG was assessed to create bisulphite plots.

2.5 Protein extraction and analysis

2.5.1 Protein extraction and quantification

For protein extraction, cell pellets were lysed in 200 µl of Pierce RIPA solution (ThermoFisher Scientific) with 1x phosphatase inhibitor solution (ThermoFisher Scientific) and 1x protease inhibitor cocktail (Sigma). Samples were pipetted multiple times to ensure cell lysis and incubated on ice for 10 minutes. 1 µl of benzonase (Sigma) was then added in order to degrade nucleic acids and samples incubated on ice for a further 30 minutes. To optimise cell lysis, samples were then sonicated in a Diagenode Bioruptor at 4 ^{o}C on high for 10 cycles at 30 seconds on/30 seconds off. To separate protein from cell debris centrifugation was done at 4 ^{o}C, 16,000 x g for 20 minutes. Supernatant was transferred to a fresh vial and stored at -20 ^{o}C.

Protein was quantified using the Pierce BCA protein assay kit according to manufacturer's instructions (ThermoFisher Scientific). Samples were diluted 1 in 5 and measured in triplicate in a 96 well flat bottomed plate. Kit standards were used to generate a standard curve. Plates were read using a Multiskan GO Microplate Spectrophotometer at 562 nm (ThermoFisher Scientific).

2.5.2 Western blotting

For western analysis, Laemmli buffer was prepared as follows:

- 1.2 g SDS
- 4.7 ml glycerol
- 1.2 ml Tris 0.5M pH 6.8
- 2.1 ml dH2O

This mixture was then rotated at 50 °C for 30 minutes before the addition of 0.93 g Dithiothretiol and 6 mg bromophenol blue

2 µl Laemmli buffer was added to a 50 µg protein sample, made up to a 10 µl solution with dH2O, and incubated at 70 °C for 3 minutes before being placed directly on ice. Protein was then loaded onto a NuPAGE 10% Bis-Tris protein gel (Invitrogen) along with PageRuler protein ladder (Invitrogen) for determination of band size. Gels were run with 1xNuPAGE MOPS SDS running buffer (Invitrogen) at 150 V for 1 hour.

After separation, the proteins were transferred onto a nitrocellulose membrane using the iBlot system (Invitrogen) using P0 template which runs at 20 V for 1 minute, 23 V for 4 minutes and 25 V for 2 minutes. Membranes were then incubated in Western Blocking Reagent (Roche) overnight at 4 °C.

Primary antibodies were diluted in block solution (α-HELLS (Proteintech 11955-1-AP) at 1:500 or α-β-TUBULIN (Abcam ab6046) at 1:1000). Membranes were incubated in primary antibody solution for 90 minutes at room temperature and then washed 3x10 minutes in wash solution (1xTBS with 0.1% Tween-20) on a rocking platform. Secondary antibodies were diluted in block solution (α-rabbit or α-mouse IgG- HRP-conjugate (Sigma) at 1:10,000). Membranes were then incubated in secondary antibody solution for 90 minutes at room temperature and then washed 3 x 10 minutes in wash solution. Membranes were then incubated with SuperSignal Western Pico Chemi Luminescence Reagent (Pierce) for 5 minutes in the dark and developed using ECL Hyperfilm (Amersham) and an X-ray processor.

2.6 RNA extraction and analysis

2.6.1 RNA extraction

For RNA extraction cell pellets were reconstituted in 1 ml TRIzol Reagent (Life technologies), mixed well by pipetting 20-30 times and incubated at room temperature for 5 minutes. 200 µl of chloroform was then added to each sample, shook vigorously and incubated at room temperature for 2 minutes. Separation of aqueous and organic phases was then achieved by centrifugation at 16,000 x g for 15 minutes at 4 °C. The aqueous phase was then transferred into a fresh vial and 500 µl of isopropanol was added. Samples were gently inverted 10 times, incubated for 10 minutes at room temperature and then centrifuged at 16,000 x g at 4 °C for 10 minutes. Supernatant was removed taking care not to disturb the pellet which was then washed in 1 ml of 70% ethanol. Samples were then centrifuged at 16,000 x g at 4 °C for 5 minutes, supernatant removed and the pellet allowed to air-dry for 5minutes. RNA was then resuspended in nuclease free water.

RNA samples underwent treatment with DNase to remove any contaminating DNA using RQ1 RNase-free DNase kit (Promega) as per manufacturer's instructions. The DNase reaction was set up with 10 µg RNA, 1 µl 10x RQ1 RNase-free DNase reaction buffer, 1 unit RQ1 RNase-free DNase made up to 10 µl reaction with dH20. Reactions were then incubated at 37 °C for 30 minutes. 1 µl of RQ1 DNase Stop Solution was added and sample incubated at 65 °C for 10 minutes to inactivate the DNase. Integrity was determined by gel electrophoresis. Intact RNA shows sharp 28S and 18S rRNA bands in a 2:1 ratio. Degraded RNA will not have these sharp bands and will have a smeared appearance. RNA was stored at -80 °C.

2.6.2 Preparation of cDNA

1 µg of total RNA was used as a template to generate complementary DNA (cDNA) using the SuperScript III kit (Invitrogen) as per manufacturer's instructions. The reaction was set up as follows:

- 1 µg of total RNA
- 1 µl Random Primers (Promega)
- 1 µl 10 mM dNTP mix (Invitrogen)
- made up to 13 µl with dH2O

Samples were incubated at 65 °C for 5 minutes and then placed on ice. To each sample was added:

- 200 units SuperScript III
- 4 µl 5x First Strand buffer
- 1 µl 0.1M DTT
- 1 µl RNaseIn (Promega)

Negative controls were also set up without the addition of the reverse transcriptase enzyme to control for the presence of genomic DNA in qRT-PCR experiments. Samples were then incubated at 25 °C for 5 minutes, 50 °C for 1 hour and the reverse transcriptase enzyme was then inactivated at 70 °C for 15 minutes. cDNA was stored at -20 °c.

2.6.3 qPCR

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to determine gene transcript levels using cDNA. Primers used were either in house or designed using the Primer-BLAST tool (Ye et al., 2012). Primers were designed to span an exon-exon junction and give an annealing temperature of 60 °C and product less than 300bp in length. Products were screened by gel electrophoresis for the formation of multiple bands or small bands indicating primer-dimer formation. Details of primers used are listed in Table 2.2. In neurodifferentiation experiments primers for lineage markers specific to each stage of

neurodifferentiation were chosen. Namely *Sox1* was selected as the earliest neuroectodermal marker, *Nestin* as a neural progenitor cell marker and *Tuj1* as a postmitotic neural marker. *Nanog* was selected as a pluripotency marker to detect undifferentiated ES cells. To detect differentiation to non-neural lineages *Gata6* was used as an endodermal marker and *Snail* as a marker of mesodermal lineage.

In chapter 5 primers were designed for selected immune response genes as follows; *Irf7* was selected as it is a key transcription factor involved in the activation of viral inducible genes. *Ccl3* and *Ccl4* were selected as chemokines which have been shown to play a role in virally induced neurological conditions such as HIV dementia (Levine et al., 2009; Olivier et al., 2018). *Casp4* was selected as it plays a role in an active inflammatory response cleaving precursors of inflammatory cytokines.

Reactions were performed in 96 well plates containing:

- 1 µl cDNA (diluted 1:25 in dH2O)
- 1x SYBR Select Master Mix (Thermo Fisher Scientific)
- 2.8 µl forward and reverse primer mix (2.5 µM of each primer)
- 3.2 µl dH2O

PCR was carried out using the LightCyler 480 System 11 (Roche). Samples were heated at 95 ^{o}C for 2 minutes and then underwent 50 cycles at 95 ^{o}C for 15 seconds and 60 ^{o}C for 1 minute.

Expression was calculated using the relative standard curve method (Cikos et al., 2007). A standard curve was generated from serial dilution of a cDNA sample from cells or tissue where the gene was suitably expressed. Results were normalised to housekeeping gene, *Gapdh* expression. Results are displayed in arbitrary expression units (AEU).

Target Gene/Primer name	**Application**	**Sequence (5'-3')**
Nanog	cDNA	F: CCTCCAGCAGATGCAAGAA R: GCTTGCACTTCATCCTTTGG
Sox1	cDNA	F: GCAGCGTTTCCGTGACTTTAT R: GGCAGAACCACAGGAAAGAAA
Nestin	cDNA	F: AGGACCAGGTGCTTGAGAGA R: TTCGAGAGATTCGAGGGAGA

Tuj	cDNA	F: AGCGATGAGCACGGCATAG R: CAGGTTCCAAGTCCACCAGA
Gata6	cDNA	F: GGTCTCCTACAGCAAGATGAATGG R: TGGCACAGGACAGTCCAAG
Snail	cDNA	F: GTCTGCACGACCTGTGGAA R: CAGGAGAATGGCTTCTCACC
Gfap	cDNA	F:CGGCGATAGTCGTTAGCTTC R:GCACTCAATACGAGGCAGTG
β-actin	cDNA	F:AGAGCTATGAGCTGCCTGACG R:TGTGTTGGCATAGAGGTCTTTACG
Gapdh	cDNA	F:GGTCCTCAGTGTAGCCCAAG R:ACCCAGAAGACTGTGGATGG
5'IAP LTR	cDNA	F:GATGGTGCTGACATCCTGTG R:CTGACGTTCACGGGAAAAAC
MAJ sat	cDNA	F:AAATACACACTTTAGGACG R:TCAAGTGGATGTTTCTCATT
Maj Sat	DNA bisulphite primers	F:GGAATATGGTAAGAAAATTGAAAATTATGG R:CCATATTCCAAATCCTTCAATATACATTTC
L1-5'UTR	cDNA	F:AATCTGTCTCCCAGGTCTGC R:CCTTTCGCCATCTGGTAATC
L1 ORF	cDNA	F:ATCCAGGAAATCCAGGAC R:TTTGCTGGACCTTTGAGTTG
HoxA5	cDNA	F:CTCATTTTGCGGTCGCTATCC R:ATCCATGCCATTGTAGCCGTA
HoxB5	cDNA	F:CCTGCACTAACGGCGACAG R:TGGCCTCGTCTATTTCGGTGA
HoxC5	cDNA	F:TCCCTGCCTATAACATGCAAAC R:CAATCCGCCGTAGCAGTACC
HoxD3	cDNA	F:CAGAAGGCTGCATACTACGAG R:GAACTTGGGTAATCACTGTCCAG
Primer for recombination efficiency	DNA	F:TGTGGATAATTCCATTTCACAG R:ATGTATGCACTAAAGAACGAC
β-actin	DNA	F:GGCTGCAAAGAGTCTACACC R:TCACTCAGAACGGACACCAT
HoxC9	DNA	F:TCAGTCTGGGCTCCAAAGTC R:AGAGGTAGCCTCCCCAGAAC
Ccl4	cDNA	F:TTCCTGCTGTTTCTCTTACACCT

		R:CTGTCTGCCTCTTTTGGTCAG
Casp4	cDNA	F:AGCGTTGGGTTTTTGTAGATGC R:CCTTGTGAACTCTTCAGGGGA
Irf7	cDNA	F:GCGTACCCTGGAAGCATTTC R:GCACAGCGGAAGTTGGTCT
Ccl3	cDNA	F:TGTACCATGACACTCTGCAAC R:CAACGATGAATTGGCGTGGAA

Table 2.2. Primer Sequences

Bisulphite primers used were taken from Dunican *et al (Dunican et al., 2013).* Primers were designed to target mouse major satellite GenBank: EF028077 at chr9: 3,017,919-3,024,581. This 6kb region is made of repetitive units with the primers targeting a sequence in these units containing 3 CpGs as highlighted in figure 2.5.

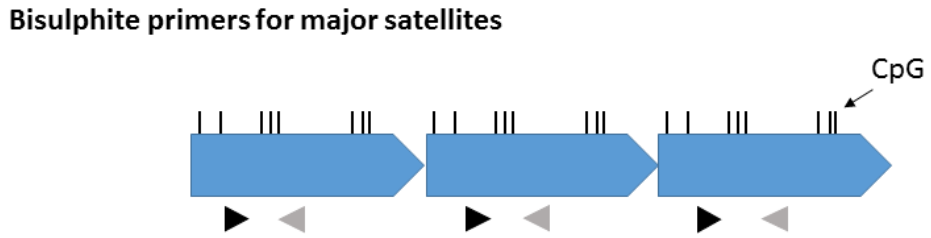


Figure 2.5. Map of bisulphite primers. Diagram displays primer positions on major satellites. Black and grey triangles indicate primer pairs. CpGs are indicated.

2.7 Next generation sequencing experiments

2.7.1 RNA-Seq

The transcriptomes of cell and mouse models were analysed using RNA-Sequencing. Before library preparation the integrity of the extracted RNA was determined using the Bioanalyzer 6000 Nano Chip (Agilent) by HGU technical services. RNA was considered fit for use if it had a RIN score ≥ 8. Three biological replicates for each condition were sequenced.

I enriched for polyadenylated mRNA and prepared sequencing libraries using the NEBNext Poly(A) mRNA Magnetic Isolation Protocol Kit (E7490), the NEBNext Ultra Directional RNA Library Prep kit for Illumina (E7420s) and NEBNext Multiplex Oligos for Illumina (E7335s) according to the manufacturer's instructions:

mRNA isolation, fragmentation and priming

mRNA was isolated by mixing total RNA with poly (T) oligomers covalently attached to magnetic beads.

20 µl Oligo d (T) 25 beads were washed twice in 100 µl 2x RNA binding buffer and placed on a magnetic rack to remove supernatant. Beads were then resuspended in 50 µl 2x RNA binding buffer and 1 µg total RNA, made up to 50 µl with nuclease free water, was added to the beads. Samples were incubated at 65 ^{o}C for 5 minutes and then cooled to 4 ^{o}C. Beads were resuspended and then incubated at room temperature for 10 minutes with gentle mixing half way through. Supernatant was removed and beads washed twice in 200 µl of wash buffer. After removal of supernatant 50 µl of Tris buffer was then added and incubated at 80 ^{o}C for 2 minutes and cooled to 25 ^{o}C. 50 µl 2x RNA binding buffer was added to the sample and it was incubated at room temperature for 5 minutes before removal of supernatant. Beads were then washed in 200 µl of wash buffer. mRNA was eluted from the beads following the addition of 6.2 µl 5x NEBNext first strand synthesis buffer, 1.55 µl NEBNext random primers and 7.75 µl nuclease free water. Samples were incubated at 94 ^{o}C for 15 minutes. 13.5 µl of eluted mRNA was transferred to a fresh PCR tube and placed on ice.

First strand cDNA synthesis

To the sample was added the following:

- 0.5 µl murine RNAse inhibitor
- 1 µl protoscript II reverse transcriptase
- 5 µl Actinomycin D (0.1 µg/ ul)

Samples were then placed in a thermal cycler at 25 ^{o}C for 10 minutes, 42 ^{o}C for 15 minutes, 70 ^{o}C for 15 minutes and then cooled to 4 ^{o}C.

Second strand cDNA synthesis

To the sample was added the following

- 8 µl 10x second strand synthesis reaction buffer
- 4 µl second strand synthesis enzyme mix
- 48 µl nuclease free water

Samples were then incubated at 16 ^{o}C for 1 hour.

Purification of ds cDNA

144 µl AMPure XP beads was then added to the reaction and incubated at room temperature for 5 minutes. Supernatant was discarded and beads washed in 200 µl 80%

ethanol. This wash step was repeated and beads allowed to air dry. DNA was eluted from the beads with the addition of 60 µl 0.1xTE and stored overnight at -20 °C.

Library Preparation

End preparation of cDNA library

To the sample was added 6.5 µl of 10x NEBNext end repair reaction buffer and 3 µl NEBNext end prep enzyme mix. Samples were then incubated at 20 °C for 30 minutes, 65 °C for 30 minutes and cooled to 4 °C.

Adaptor ligation

To each sample was then added 15 µl blunt/TA ligase master mix and 1 µl NEBNext adaptor diluted 1:9 with nuclease free water and 2.5 µl nuclease free water. Solution was incubated at 20 °C for 15 minutes. 3 µl of USER enzyme was then added and incubated at 37 °C for 15 minutes.

Purification of ligation reaction

Each reaction was made up to 100 µl with nuclease free water and 100 µl of AMPure XP beads added. Samples were then incubated at room temperature for 5 minutes. Samples were placed on a magnetic rack, supernatant discarded, and beads washed in 200 µl 80% ethanol. This wash step was repeated and samples then air dried. DNA was eluted from beads with 52 µl 0.1 X TE and transferred to a clean tube. To this was added 50 µl AMPure XP beads and the whole process repeated with a final elution volume of 20 µl.

PCR enrichment

To the sample was added 2.5 µl NEBNext Q5 hot start HIFI PCR master mix, 2.5 µl universal primer and 2.5 µl of index primer. Samples then went into a thermal cycler with the following cycling conditions: 98 °C for 30 seconds followed by 13 cycles of 98 °C for 10 seconds, 65 °C for 75 seconds followed by 65 °C for 5 minutes.

Purification of PCR reaction

45 µl AMPure XP beads were then added to each 50 µl PCR reaction and incubated at room temperature for 5 minutes. Tubes were then placed on a magnetic rack and supernatant removed. Beads were then washed in 200 µl of 80% ethanol and beads air dried. DNA was the eluted in 20 µl 0.1xTE and stored at -20 °C.

Resulting libraries were quality checked and quantified by HGU technical services using an Agilent Bioanalyser high sensitivity chip. Samples were then pooled and sequenced by Edinburgh Genomics using Illumina HiSeq 4000 75bp paired-end reads.

2.7.2 ERRBS

Enhanced reduced representation bisulphite sequencing was employed to generate genome-wide DNA methylation profiles. 250 ng of genomic DNA was sent to the Epigenomics Core Facility of Weill Cornell Medicine, New York where libraries were generated using their in-house protocol and sequenced using the Illumina CASAVA 1.8.2 pipeline (Garrett-Bakelman et al., 2015). One replicate for each condition was sent.

2.7.3 ChIP-seq

Cross-linked chromatin immunoprecipitation followed by sequencing was utilised to map H3K27me3 genome-wide. This is done by chemically cross-linking proteins bound to the DNA with formaldehyde. The cells are then lysed and chromatin fragmented by sonication and an antibody against H3K27me3 is then used to pull down the regions of bound DNA. Solutions used in experiments are detailed in Table 2.3.

LB1	**50 mM Hepes KOH pH7.9** **140 mM NaCl** **1 mM EDTA** **10% glycerol** **0.5% IPEGAL** **0.25% Triton X-100** **1xComplete EDTA-free protease inhibitor tablet (Roche)**
LB2	10 mM TrisHcl pH8 200 mM NaCl 1 mM EDTA 0.5 mM EGTA 1xComplete EDTA-free protease inhibitor tablet (Roche)
LB3	10 mM TrisHCl Ph8 100 mM NaCl 1 mM EDTA 0.5 mM EGTA 0.1% Na Deoxycholate 0.5% LaurylSulphate 1xComplete EDTA-free protease inhibitor tablet (Roche)
Chip Dilution Buffer	1% Triton X-100100

	1 mM EDTA 20 mM TrisHcl pH8 150 mM NaCl 1xComplete EDTA-free protease inhibitor tablet (Roche)
Block Solution	1xPBS 0.5% BSA 1xComplete EDTA-free protease inhibitor tablet (Roche)
Low salt buffer	0.1% SDS 1% Triton X-100100 2 mM EDTA 20 mM TrisHcl pH8 150 mM NaCl
High salt buffer	0.1%SDS 1% Triton X-100100 2 mM EDTA 20 mM TrisHcl pH8 500 mM NaCl
LiCl buffer	0.25 M LiCl 1% IPEGAL 1% NaDeoxycholate 1 mM EDTA 10 mMTrisHCl pH8
Elution Buffer	0.1 M NaHCO3 1% SDS

Table 2.3. **Solutions used in ChIP experiments.**

For each sample $9x10^6$ cells were harvested by trypsinisation and rinsed well in 1xPBS before being resuspended in 1.5 ml of PBS. Each sample was fixed with the addition of 40.5 µl of 37% formaldehyde (Sigma) and incubated for 10 minutes at room temperature on rotation. 225 µl of 1M glycine was added to quench the formaldehyde and rotated at room temperature for 5 minutes. Samples were then centrifuged at 300 x g for 4 minutes at 4 ^oC, supernatant discarded and pellets resuspended in 2 ml LB1 solution and rotated at 4 ^oC for 4 minutes. This process was repeated resuspending in 2 ml LB2 solution and then repeated resuspending in 200 µl LB3.

Chromatin was then sheared at 4 ^oC using a Diagenode bioruptor set on high for 45 cycles at 30seconds on/ 30seconds off. 20 µl of 10% Triton X-100 in LB3 was then added to each

sample and mixed by inversion. Samples were then centrifuged at 16,000 x g at 4 °C for 10 minutes to remove nuclear debris and supernatant transferred to a fresh tube.

In order to check the efficiency of the sonication process, 8 µl of sheared chromatin was taken, mixed with 1 µl 5M NaCl and 16 µl dH2O, incubated at 98 °C for 15 minutes and treated with RNase cocktail before being purified with Qiagen PCR Clean up kit as per manufacturer's instructions . Sheared chromatin was then resolved on a 1.5% agarose gel post-stained with EtBr. The DNA smear should be between 100-500bp with a peak at 250-300bp.

For immunoprecipitation, 25 µl of Protein G Dynabeads (Novex Lifetech) were added to fresh tubes and washed 3 times in block solution and then resuspended in 500 µl of block. Beads were then incubated with 5 µl of antibody (H3K27me3 (Millipore 07449), IgG (Santa Cruz Biotech sc-2027)) and incubated at 4 °C for 4 hours on a rotator. Beads were then washed 3 times in 500 µl block solution and supernatant removed.

210 µl of sheared chromatin was mixed with 1890 µl of chip dilution buffer. 10% of this mix was stored at 4 °C and labelled as "input" samples. 1ml of the solution was added to the antibody bound beads and incubated at 4 °C on rotation overnight.

Supernatant was then removed from beads and they were washed at 4 °C in a series of steps with 1ml low salt buffer, 1ml high salt buffer, 1ml LiCl buffer, 1ml TE and again with 1ml TE. Beads were then resuspended in 100 µl elution buffer before being shook at maximum rpm in a thermomixer at room temperature for 30 minutes. Supernatant was then transferred to fresh tubes. To these samples, and to the input samples was added 4 µl 5M NaCl and samples incubated at 65 °C for 3 hours. 2 µl RNase cocktail was then added and reactions incubated at 65 °C for 90 minutes. 1 µl proteinase k 20 mg/ml was added and reactions incubated at 45 °C for 1 hour. Samples were then purified using Qiagen minelute kit as per manufacturer's instructions.

DNA was quantified using Quibit and 3 ng of DNA was then used for preparation of ChIP-Seq libraries using NEBNExt Ultra DNA Library Prep kit for Illumina (E7370) and NEBNExt Multiplex Oligos for Illumina (E7335) as previously described in section 2.7.1.

Sample quality check and quantification was carried out using an Agilent Bioanalyser high sensitivity chip by HGU technical services. Samples were pooled and sequenced by Edinburgh Genomics using Illumina HiSeq 4000 75bp paired-end reads. One replicate for each condition was sequenced.

2.8 Bioinformatic analyses and statistics

For the majority of analyses described in this section the R programming language (http://www.r-project.org/) version 3.4.4 and its associated packages (https://cran.r-project.org/) were used.

Where sample size was large and one could not assume a normal distribution the non-parametric Wilcoxon-Rank-Sum test in R was used to test the null hypothesis. For comparisons between a small number of samples or when normality could be assumed a Student T-test in Microsoft Excel was used to test the null hypothesis. Bar plots and line graphs were generated in Excel. Scatter plots, boxplots, violin plots and volcano plots were generated in R. Hypergeometric testing of overlapping gene lists was performed using R phyper. For comparisons between bisulphite samples the exact binomial test in R was used.

Gene ontology analysis was performed using the biological processes tool of DAVID version 6.8 (https://david.ncifcrf.gov/). This programme uses the Fisher Exact test to measure the gene-enrichment in annotation terms between a user supplied gene list and the genome background. The online tool, Interferome version 2.01 (http://www.interferome.org/) was used to search for interferon responsive genes within gene lists.

The Mus musculus mm9 build from the UCSC browser was used to map all reads from next generation sequencing experiments and genomic annotations were taken from this build by downloading the RefGene and RepeatMasker tables (http://genome.ucsc.edu/cgi-bin/hgTables).

I am very grateful to Dr Donncha Dunican for carrying out all bioinformatic processing and analysis from the next generation sequencing experiments described below. For all experiments, paired-end sequence read sanfastq output files returned from Edinburgh Genomics and Weill Cornell Medicine were used. The FastQC V0.11.4 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was then used to test for sequence read quality. Trim Galore! V0.4.1 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was then used to remove adaptor sequences and sequences with PHRED quality scores <20. PCA plots from RNA-seq and CHIP-seq experiments are in Appendix Figure A4. These demonstrate that variance is due mainly to cell type and secondarily to genotype. They also reveal no gross outliers between biological replicates.

2.8.1 RNA-Seq Analysis

For RNA-Seq experiments, results are from three biological replicates. After initial bioinformatic processing described above, paired reads were then aligned to the mm9 genome using TopHat V2.1.0 (http://tophat.cbcb.umd.edu/). The HTSeq count tool V0.71

(https://htseq.readthedocs.io/) was then used to determine raw counts over mapped sequences and converted into an EdgeR object. The EdgeR programme V 3.12.1 (http://www.bioconductor.org/packages/2.11/bioc/html/edgeR.html) was then used to normalise read counts and determine differential gene expression. Genes with a read depth < 10 were filtered out. Transcripts were considered differentially expressed if they had a greater than 2 fold change and P value <0.05.

For repeat data only those reads with a MAPQ score of 50 were included in the analysis indicating the highest probablilty of uniquely mapped reads. Repeats with read depth <10 were filtered out. Repeats were considered differentially expressed if they had a greater than 4 fold change and FDR <0.05. Repeat expression was calculated under two different conditions. The first analysing reads mapping to full length annotated repeats i.e. “long” repeats defined as >6kb for LINE1s and LTRs and >1kb for satellites and the second analysing those mapping to non-full length i.e. “short” repeats.

2.8.2 ERRBS Analysis

Following initial processing as described above, ERRBS data was mapped to the mm9 genome using the Bismark tool V0.16.3 (http://www.bioinformatics.babraham.ac.uk/projects/bismark). The “bismark_methylation_extractor” script was used to extract methylation scores over individual CpGs. CpGs with a read depth <5 were filtered out. To determine methylation over promoter regions the average methylation across all CpGs spanning the +/- 2kb region across the TSS was calculated. For repeats, average methylation of CpGs mapping to long repeats of each class was calculated. A region was considered differentially methylated if it had a greater than 20% absolute methylation difference.

2.8.3 ChIP-seq Analysis

Following initial bioinformatic processing, ChIP data was aligned using Bowtie2 (http://bowtie-bio.sourceforge.net/index.shtml). Again only reads with MAPQ score of 50 were included in the analysis. HTSeq count tool was then used to determine raw counts. An in-house pipeline was then used. Data was normalised to library size. To calculate H3K27me3 over promoter regions, reads from 1kb windows, sliding by 200bp, across the +/- 2kb region from the TSS were averaged. H3K27me3 across repeats was calculated as the number of reads mapping to an annotated “long” repeat. Regions with read depth < 10 were filtered out. Regions were considered to be differentially marked for H3K27me3 if there was a greater than 2 fold difference in reads.

Chapter 3. *Lsh* is required for DNA methylation during neurogenesis

3.1 Introduction

The requirement for *Lsh* to establish normal DNA methylation has been well recognised. Despite the fact that *Lsh* is highly expressed in the brain during development, there is an absence of research into its role specifically in the nervous system. Indeed, at the beginning of this project it was unknown if *Lsh* was required for appropriate DNA methylation during neurogenesis and if its presence was essential for normal neurogenesis. In this chapter I aim to address these questions, the importance of which became highly relevant with the recent discovery of Type 4 ICF syndrome caused by mutations in the human homolog of *Lsh* (Alghamdi et al., 2018; Thijssen et al., 2015).

Early embryonic development can be modelled *in vitro* using embryonic stem cells derived from the ICM of the blastocyst. These are a pluripotent cell population that can then be induced to differentiate down discrete lineages. In order to demonstrate the requirement for *Lsh* during neurodevelopment I firstly optimised a cell differentiation system using N2B27 media adapted from Pollard *et al* (Pollard et al., 2006) which favours neural differentiation. Traditional methods of neural cell derivation from ES cells have relied on generation of multicellular aggregates, which are then exposed to retinoic acid or grown in conditioned media, or co-culture with feeder cells (Bain et al., 1995; Kawasaki et al., 2000). However, the method I have adopted is a monolayer serum-free protocol. There are advantages to this as it relies on an autocrine induction mechanism where FGF signalling occurs in the absence of BMP4 signalling (Abranches et al., 2009). It therefore does not rely on unknown media components with unknown physiological relevance, or retinoic acid, which although an efficient inducer of neural differentiation, also induces neural progenitors to develop a more restricted posterior fate (Wichterle et al., 2002). Furthermore this rapid protocol allows direct observation of cell morphology and results in a higher yield of neural cells compared to other protocols (Cai and Grabel, 2007).

I utilised this system to firstly determine whether ES cells lacking *Lsh* could differentiate normally down neural lineage. Other cell lines knocked out for key regulators of the methylation process have shown defects in this. For example, *Dnmt1* KO ES cells undergo apoptosis when induced to differentiate (Jackson et al., 2004; Lei et al., 1996) and *Dnmt1* and *Dnmt3a* KO neural progenitor cells have accelerated glial differentiation at the expense of neuronal (Fan et al., 2005a; Wu et al., 2010; Wu et al., 2012). Secondly, I looked at the

requirement for *Lsh* in global DNA methylation in this system, and finally the effects of rescuing *Lsh* expression in the knockout cells.

3.2 Results

3.2.1 Characterisation of *Lsh*$^{-/-}$ mES cells

Lsh knockout mES cells were generated by Dr Donncha Dunican using the CRISPR/Cas9 gene editing system. E14 mES cells were transfected with Cas9 and a single guide RNA designed to target exon 1 of the *Lsh* gene. The clone I have used was shown to have a homozygous 25bp deletion in exon 1. I confirmed this line does not express LSH at the protein level via western blot (Figure 3.1). This line was used for all analyses, with wild type E14 mES cells (WT) as a control, henceforth this knockout cell line will be referred to as *Lsh*$^{-/-}$.

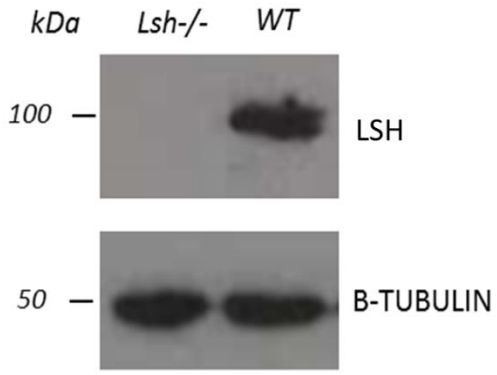


Figure 3.1. *Lsh*$^{-/-}$ mES cells do not express LSH at the protein level. Confirmation that *Lsh*$^{-/-}$ mES cells do not express LSH at the protein level is demonstrated by western blot analysis of protein extracts derived from WT and *Lsh*$^{-/-}$ mES cells using α-HELLS (human homologue of LSH) antibody and α- β-TUBULIN antibody as a loading control.

The *Lsh*$^{-/-}$ cells displayed a slightly reduced growth rate, but similar morphology to WT cells (Figure 3.2 A & B), and importantly demonstrated comparable expression of the pluripotency marker *Nanog* as assessed by immunocytochemistry (Figure 3.2 C) and later by qRT-PCR (Figure 3.4 C). It should be noted that towards the end of this project an off-target mutation was identified in the *Lsh*$^{-/-}$ cells by Dr Ailsa Revuelta (see Figure A1 Appendix).

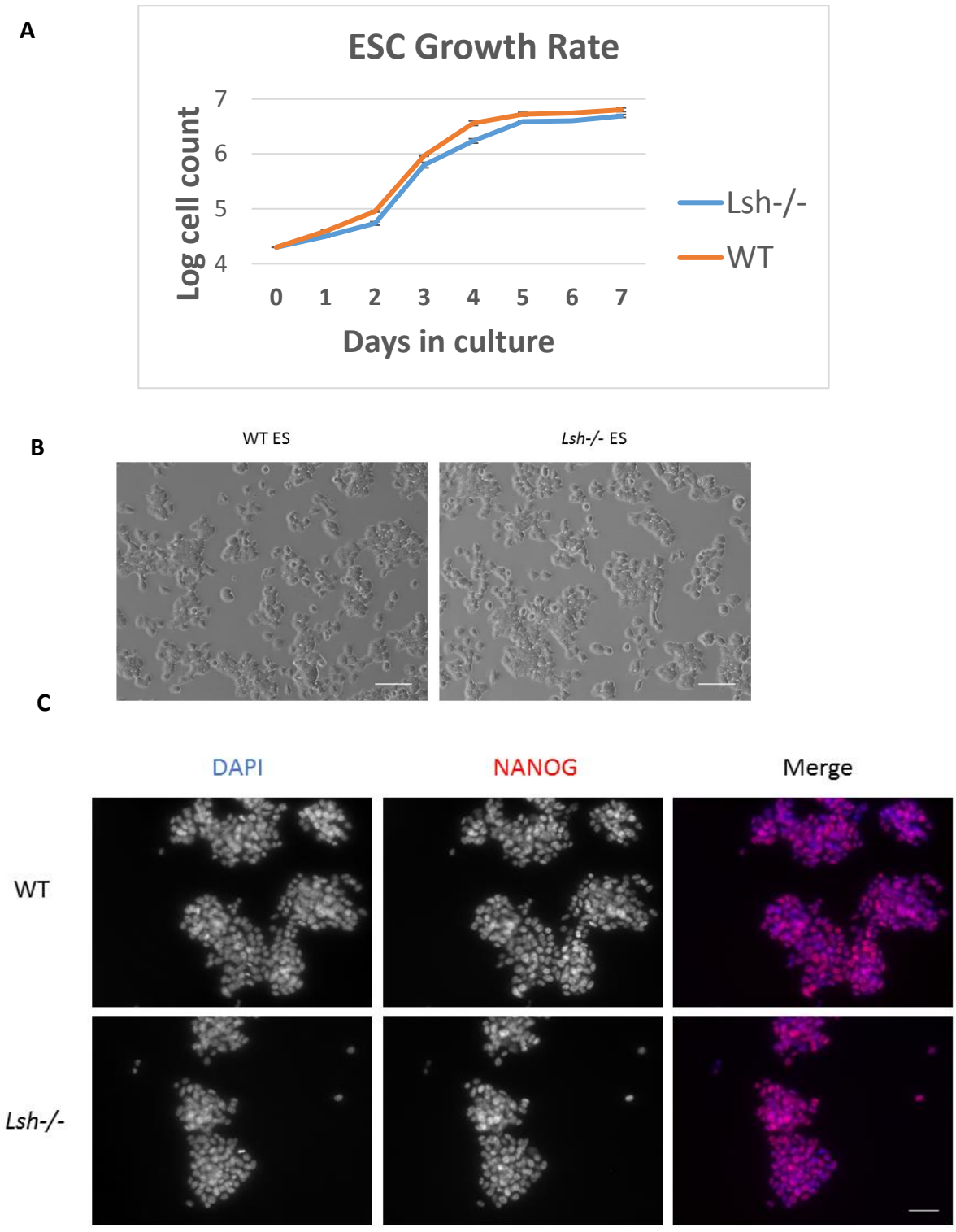


Figure 3.2. Characterisation of *Lsh*$^{-/-}$ mES cells. A. Growth rates of *Lsh*$^{-/-}$ and WT cells maintained in serum culture. Error bars represent the SEM of 3 replicates. **B.** WT and *Lsh*$^{-/-}$ ESCs show similar morphology. Scale bars represent 100 µm. **C.** Immunofluorescent images of WT and *Lsh*$^{-/-}$ ES cells using antibodies against the pluripotency marker NANOG. Nuclear DNA was stained using DAPI. Scale bars represent 100 µm.

3.2.2 *Lsh*$^{-/-}$ cells show greater propensity for differentiation down neural lineage

In order to induce neural differentiation, WT and *Lsh*$^{-/-}$ ES cells were then put through an N2B27 differentiation protocol. To confirm differentiation and compare neurogenesis between the cell lines, cultures were examined for morphological changes and neural marker gene expression.

As *Lsh*$^{-/-}$ MEFS display a replicative senescent phenotype accumulating high centrosome numbers, micronuclei formation and multipolar spindles (Fan et al., 2003), I examined metaphase spreads during neurodifferentiation of *Lsh*$^{-/-}$ cells however, I did not detect any accumulation of chromosomal abnormalities (Appendix Figure A2).

3.2.2.1 WT and *Lsh*$^{-/-}$ cells develop neuronal cell morphology during N2B27 differentiation

An advantage to using a monolayer neurodifferentiation protocol is the ability to clearly observe changes in morphology allowing quick assessment of developmental stage and comparison between different cell lines. *Lsh*$^{-/-}$ and WT mESCs were grown in N2B27 differentiation media for a total of 10 days. During this time they develop typical neuronal morphology with the appearance of long axons indicating mature post-mitotic neurons (Figure 3.3). Consistently across 5 separate differentiation experiments, a larger proportion of cells displayed mature neuron morphology at an earlier stage (by day 7) in *Lsh*$^{-/-}$ cultures.

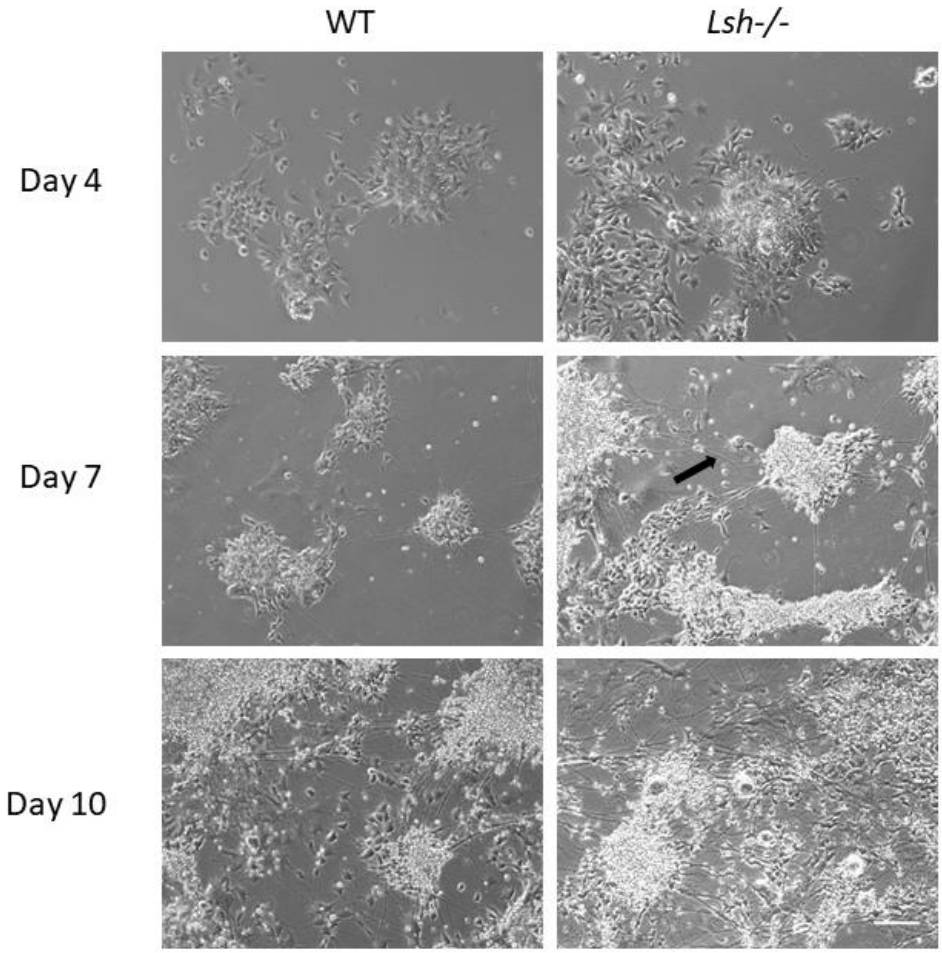


Figure 3.3. WT and *Lsh*$^{-/-}$ cells develop neuronal morphology during N2B27 differentiation. Images of WT and *Lsh*$^{-/-}$ cells during neural differentiation at days 4, 7 and 10 of protocol. The black arrow indicates a neural axon. Scale bars represent 100 µm.

3.2.2.2 Neural genes become expressed during N2B27 differentiation

To better visualise and quantify neurodifferentiation, I carried out immunocytochemistry and qRT-PCR to evaluate expression of neural lineage markers (Figure 3.4 A-C). qRT-PCR experiments were carried out on RNA extracted from 3 separate differentiation experiments.

Nanog is a transcription factor critical to self-renewal of undifferentiated ES cells and can therefore be used as a marker of pluripotency. Both WT and *Lsh*$^{-/-}$ ES cells highly expressed *Nanog* at comparable levels, and successfully down-regulated this upon differentiation (Figure 3.4 C).

Expression of *Sox1*, the earliest known specific marker of neuroectodermal lineage was evident at day 4 in both cell lines and later down-regulated to coincide with the up-regulation of more mature neural markers (Figure 3.4 C).

The expression of *Nestin* by day 4 indicated the presence of neural progenitor cells. This staining occurred at the periphery of the neural rosettes which is where one would expect to see the developing neurons (Figure 3.4 A). Significantly higher expression of *Nestin* was found in *Lsh*$^{-/-}$ cultures at days 7 and 10 of differentiation (Figure 3.4 C).

Neuron-specific class III β-tubulin gene (*Tuj1*) is a marker of post-mitotic neurons and was vastly up-regulated upon differentiation with visible staining by day 4 (Figure 3.4 B). Expression of *Tuj1* was significantly higher in *Lsh*$^{-/-}$ line by day 7(Figure 3.4 C).

GFAP, an astrocytic cell marker, was not greatly enriched in the cultures even by day 10 (Figure 3.4 C). This is in keeping with the fact that glial differentiation occurs at a later stage in neurodevelopment, both *in vivo* and under the differentiation protocol used here (Abranches et al., 2009; Qian et al., 2000; Sauvageot and Stiles, 2002; Ying et al., 2003b). However, hypomethylated cell lines such as *Dnmt1*$^{-/-}$ and *Dnmt3a*$^{-/-}$ have shown precocious astrocyte differentiation (Fan et al., 2005a; Wu et al., 2012). The developmentally regulated demethylation of glial marker genes controls the timing of astrocyte differentiation. It has therefore been proposed that existing hypomethylation of these genes, along with hypomethylation of genes encoding components of the JAK-STAT pathway, triggers this precocious astrocyte differentiation. I did not find evidence of precocious glial differentiation in the *Lsh*$^{-/-}$ cell line suggesting *Lsh* does not play a role in methylation at these sites.

Quantification of non-neural lineage marker gene expression namely *Gata6* (an endodermal lineage marker) and *Snail* (a mesodermal lineage marker) revealed some up-regulation upon differentiation but to a lesser degree than neural lineage markers suggesting the presence of non-neural cells within my cultures (Figure 3.4 C). This up-regulation was significantly greater in the WT cells. This fact, along with the significantly higher expression of neural markers in *Lsh*$^{-/-}$ cells, suggests that these knockout cells have a greater propensity to differentiate towards neural fate.

Cells were harvested at day 10 given the high expression of the neural markers *Nestin* and *Tuj* at this point and low expression of glial, pluripotent and neural stem cell markers.

Figure 3.4 A

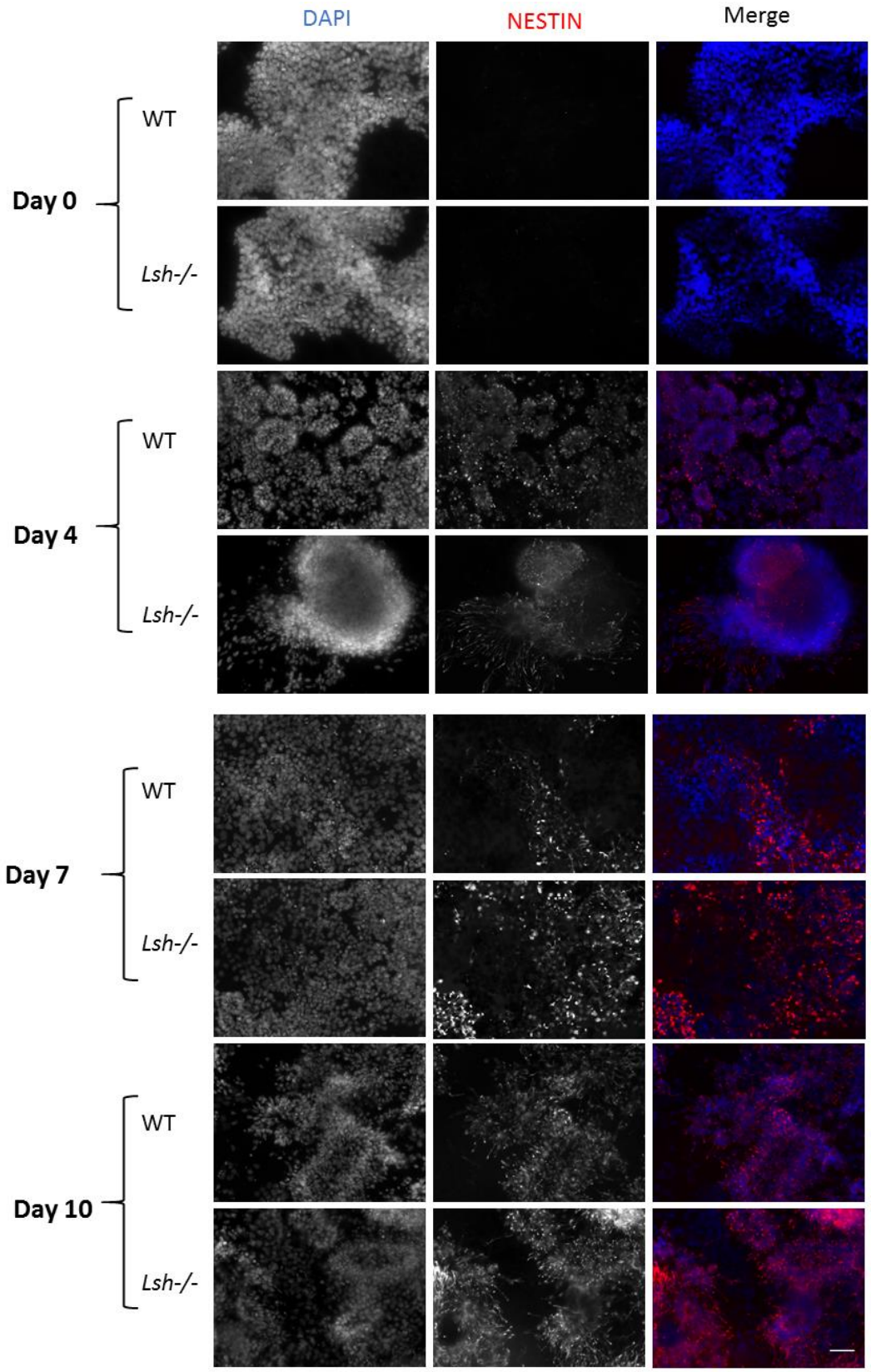
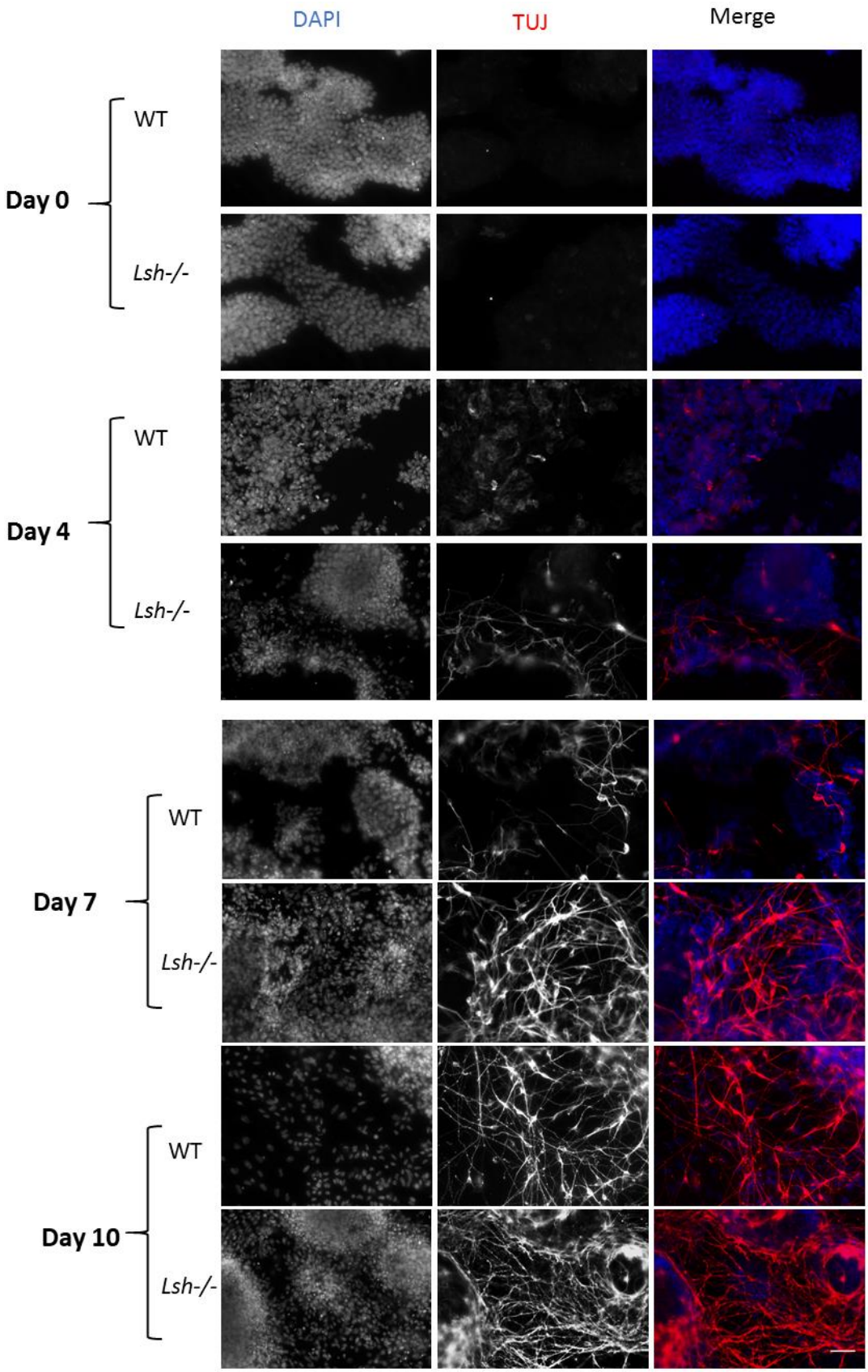

DAPI
NESTIN
Merge
Day 0
WT
Lsh-/-
Day 4
WT
Lsh-/-
Day 7
WT
Lsh-/-
Day 10
WT
Lsh-/-

Figure 3.4 B



C

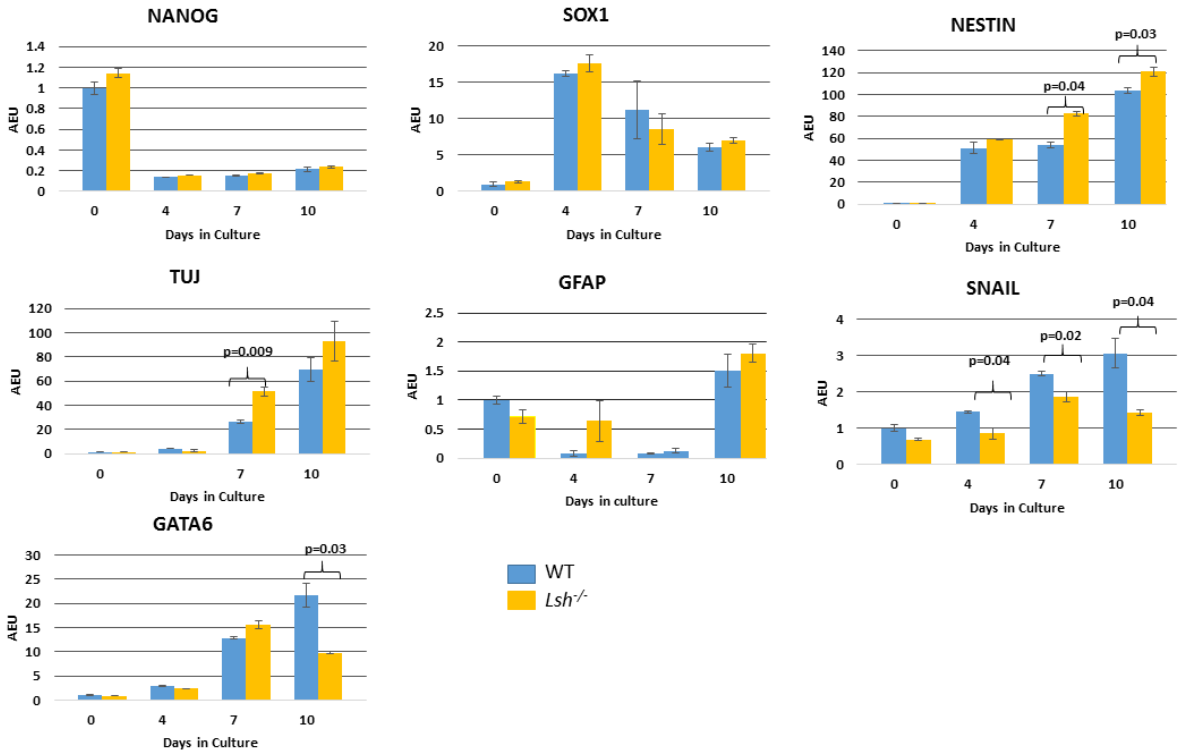


Figure 3.4. Expression of neural lineage markers during N2B27 differentiation. Immunofluorescent images of WT and *Lsh*$^{-/-}$ cells at days 4, 7 and 10 in the differentiation protocol using antibodies against NESTIN (**A**) and TUJ1 (**B**). Nuclear DNA was stained using DAPI. Scale bars represent 100 µm. **C.** Quantification of expression of lineage marker genes by qRT-PCR. Expression is normalised to housekeeping *Gapdh* expression and then standardized to WT ES expression. Error bars represent +/- SEM of 3 biological replicates. P-values indicate level of significance calculated by two-tailed student t-test. Only significant values are displayed. AEU=arbitrary expression units.

3.2.2.3 MAC sorting successfully enriches for NPCs

Given the heterogeneity between the day 10 cultures, it was essential to purify for neural progenitor cells prior to any further investigation. To do this, I employed Magnetic Activated Cell Sorting (MACS). Day 10 cell suspension was incubated with magnetic beads coated with PSA-NCAM, an NPC cell surface marker, and passed through a magnetic column. Cells that remained within the column were eluted and collected. In order to check the efficiency of this, I incubated day 10, and post MACS sorted cells with a fluorescently tagged antibody to PSA-NCAM and analysed the percentage of tagged cells via flow cytometry with the aid of Elisabeth Freyer (IGMM FACS service). This was done with cells from 3 separate differentiation experiments. This demonstrated that prior to MAC sorting, the percentage of PSA-NCAM positive cells in WT and *Lsh*$^{-/-}$ cell lines was 66.7% vs 78.6% respectively and 92.7% vs 96% following sorting (Figure 3.5). Not only does this confirm the efficiency of this method for enriching for NPCs but it also provides further evidence that the *Lsh*$^{-/-}$ cells are differentiating preferentially down neural lineage in comparison to WT. MAC sorted day 10 cells will herein be referred to as NPCs.

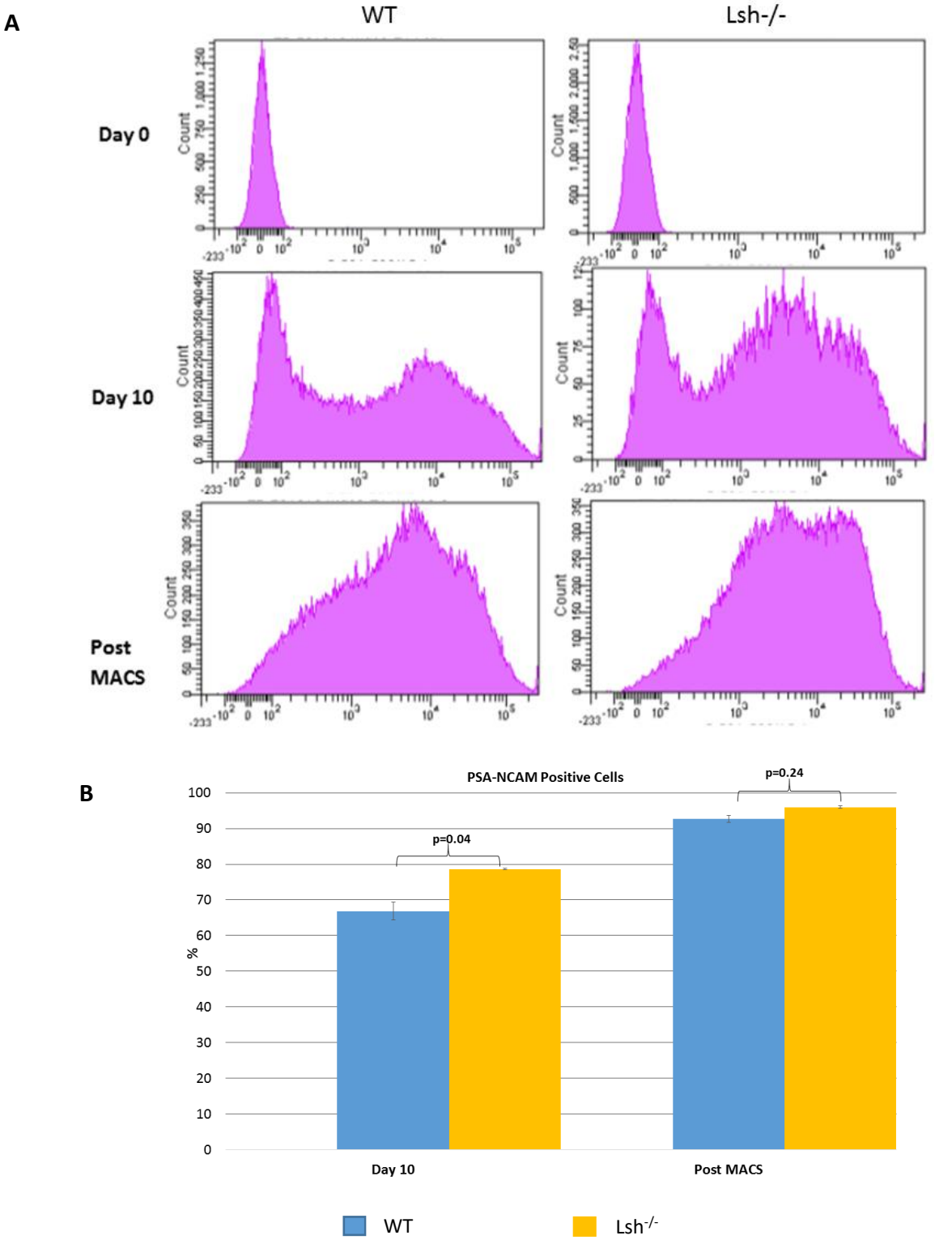


Figure 3.5. MAC sorting enriches for NPCs. A. Representative histograms, generated by flow cytometry, of PE fluorescence of cells following incubation with PSA-NCAM antibody conjugated to PE, at day 0 and 10 of differentiation and post MAC sorting. **B.** Box plot showing percentage PSA-NCAM positive cells following FACS analysis. Error bars represent +/- SEM of 3 biological replicates. P-values indicate level of significance calculated by two-tailed student t-test.

3.2.3 *Lsh* is required for DNA methylation upon neural differentiation

To assess the contribution of *Lsh* to DNA methylation upon neural differentiation, I assessed global DNA methylation levels of *Lsh*$^{-/-}$ and WT ES and NP cells. For initial assessment of this I carried out restriction enzyme digests of genomic DNA with the methylation-sensitive restriction endonuclease HpyCH4IV, which cuts at unmethylated ACGT sites which occur preferentially at satellite DNA. Given the abundance of normally methylated satellite sequences in the mouse genome, I reasoned that this would provide an indication of global DNA methylation levels. Following digestion, the DNA was separated by agarose gel electrophoresis. I observed no difference in the ability of HpyCH4IV to digest DNA from either cell line when maintained as ES cells, suggesting no difference in DNA methylation levels. However, following differentiation I observed increased digestion, indicated by increased smearing and the appearance of lower molecular weight fragments, in the *Lsh*$^{-/-}$ NPCs demonstrating hypomethylation at satellite regions (Figure 3.6 A). Digests were also carried out using the methylation-insensitive restriction endonuclease MspI, which cuts at CCGG sites regardless of methylation status, as a control for the digestibility of the DNA.

In order to quantify these global DNA methylation changes, I then digested DNA samples to single nucleotides and looked at the proportion of methylated cytosines as a percentage of total cytosines by liquid-chromatography mass spectrometry (LC-MS) performed by Jimi Wills. This method demonstrated no difference in levels of global DNA methylation in WT and *Lsh*$^{-/-}$ ES cells at 4.3% vs 4.2% respectively. However, following neural differentiation, *Lsh*$^{-/-}$ cells were significantly hypomethylated with 5mC levels reaching only 68.7% of WT levels (Figure 3.6 B).

Lsh therefore plays a key role in DNA methylation during neural differentiation. An unexpected finding was the loss of DNA methylation in the *Lsh*$^{-/-}$ cell line upon differentiation, with 5mC levels reducing from 4.2% in ES cells to 3.4% in NPCs. This is in contrast to WT cells, which show an increase from 4.3% to 4.95% as one would expect given the wave of *de novo* methylation that begins in the implanting blastocyst and continues throughout development. This suggests a previously less well described role for *Lsh* in maintenance of DNA methylation. Another possibility is that this is the cumulative effect of further demethylation / remethylation events occurring during this developmental window.

Figure 3.6 A

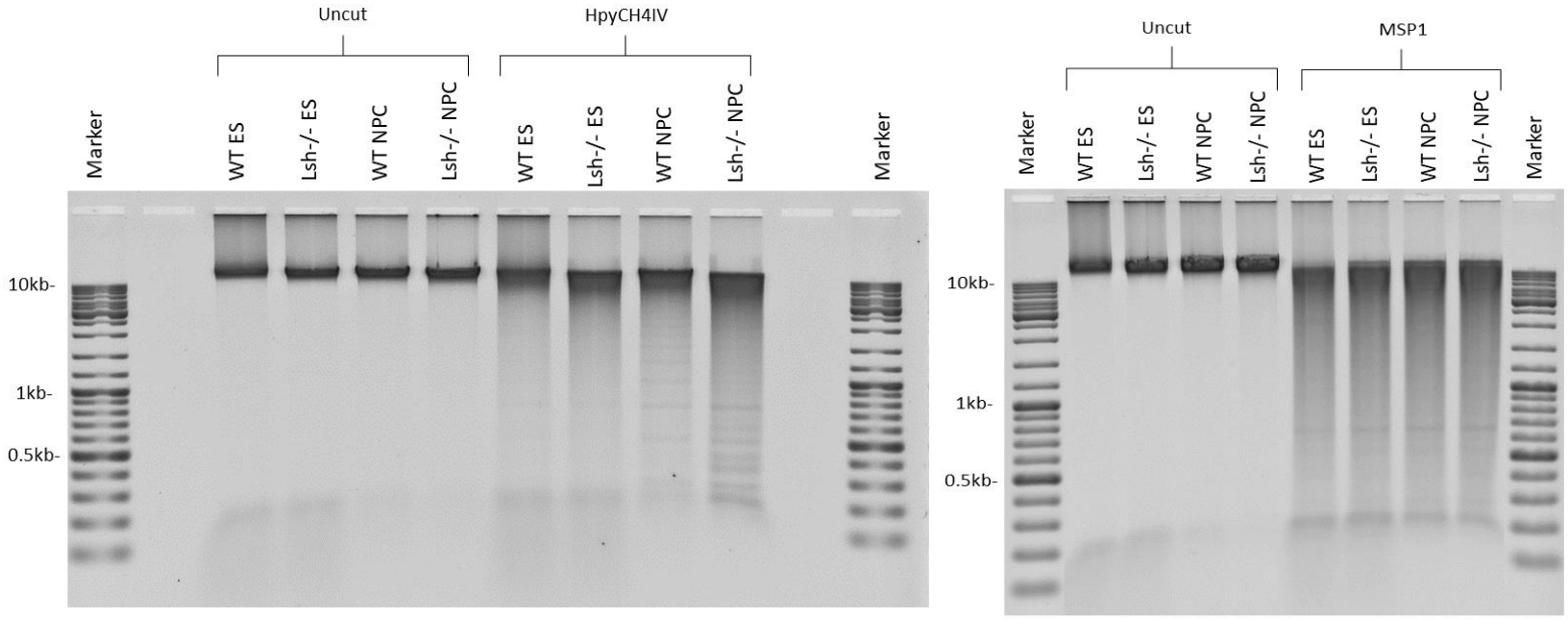


Figure 3.6. Global DNA hypomethylation in *Lsh*$^{-/-}$ NPCs. A. Agarose gels displaying digestion of DNA from WT and *Lsh*$^{-/-}$ ES and NP cells by methylation sensitive HpyCH4IV and methylation insensitive MSP1 endonucleases. The presence of low molecular weight bands indicate hypomethylation of satellite sequences. Integrity and digestibility are controlled for using uncut samples and MSP1 digestion respectively.

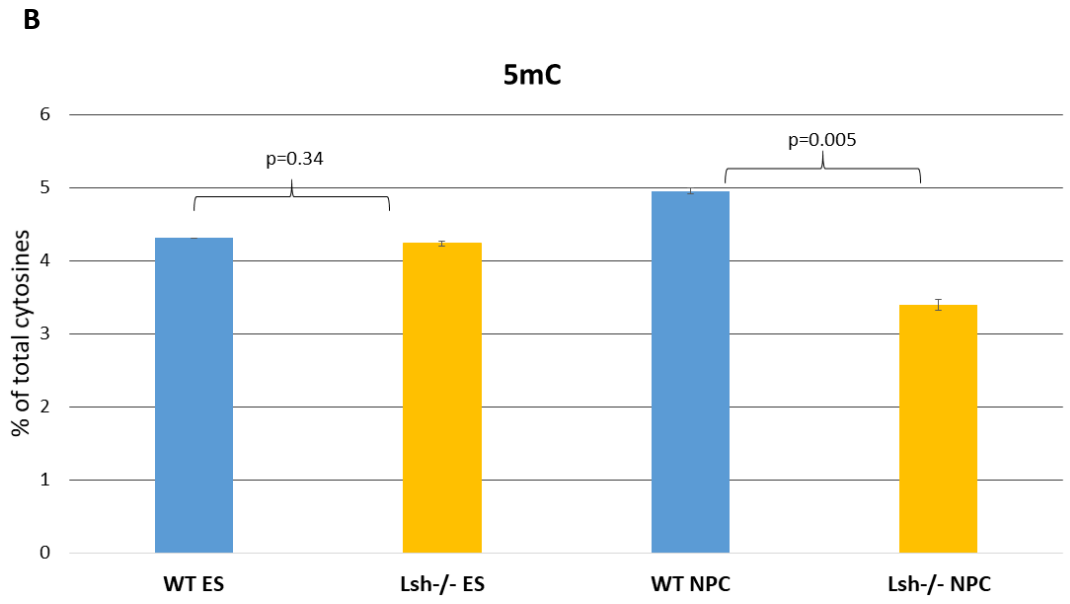


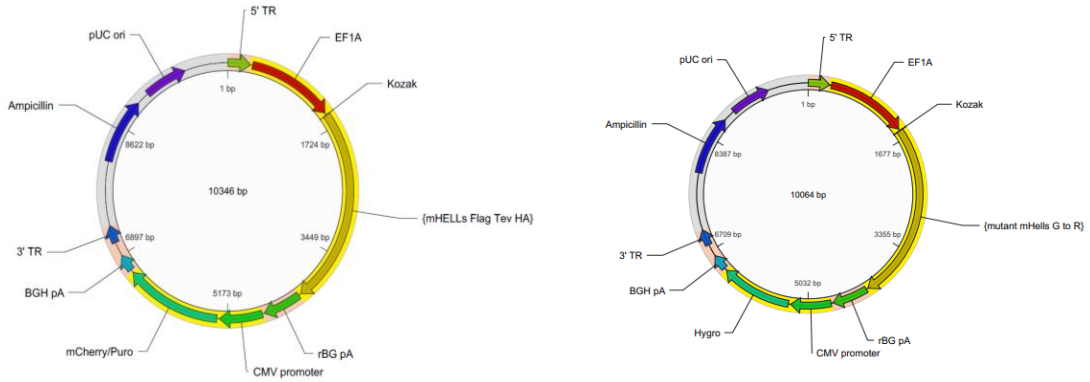
Figure 3.6. Global DNA hypomethylation in *Lsh*$^{-/-}$ NPCs. B. Quantification of global 5-mC levels by LC-MS given as a percentage of total cytosines in WT and *Lsh*$^{-/-}$ ES and NP cells. Error bars represent the S.E.M of 3 biological replicates. P-values indicate level of significance calculated by two-tailed student t-test.

3.2.4 Re-Expression of *Lsh* rescues the DNA methylation defect

3.2.4.1 Generation of *Lsh* rescue cell lines

With the purpose of confirming that the hypomethylation seen in the knockout cells is due to the absence of *Lsh,* and is not a peculiarity of the clone generated by CRISPR, or indeed a consequence of off-site targeting, I rescued *Lsh* expression in the knockout cell line by transfecting with a *PiggyBac* vector carrying wild type *Lsh* (Figure 3.7 A). A second cell line was also generated, transfecting the $Lsh^{-/-}$ cell line with a *PiggyBac* vector carrying a mutant version of *Lsh,* with a glutamine to arginine residue change at position 699 in a conserved helicase domain (Figure 3.7 B). This mutation is present in 2 patients with ICF syndrome (Thijssen et al., 2015). The *PiggyBac* transposon is a mobile genetic element that can transpose via a "cut and paste" mechanism where the *PiggyBac* transposase recognises a pair of terminal repeat sequences flanking the construct DNA and can then insert this construct DNA into random TTAA chromosomal sites in the host DNA. Thereby it is an efficient method in which multiple copies of exogenous DNA can be inserted into the genome (Rostovskaya et al., 2012). An mCherry/ Puromycin resistance dual reporter gene was inserted into the wild type rescue vector and a hygromycin resistance reporter gene into the mutant rescue vector to be used as selectable markers. Following transfection, cells were grown in media containing the appropriate antibiotic and surviving cells were then single cell sorted and clones expanded. For the WT rescue vector, mCherry positive cells were sorted via FACS prior to this. Clones were shown to express LSH at the protein level by western blot (Figure 3.7 C & D). Clone "T19" of the rescue cell line and clone "FZ" of the mutant rescue line were chosen for further experiments as the clones with LSH expression levels most comparable to WT. Henceforth these cell lines will be referred to as $Lsh^{-/-(+WT\ LSH)}$ and $Lsh^{-/-(+Mutant\ Lsh)}$ respectively.

A **B**



Courtesy of VectorBuilder

C **D**

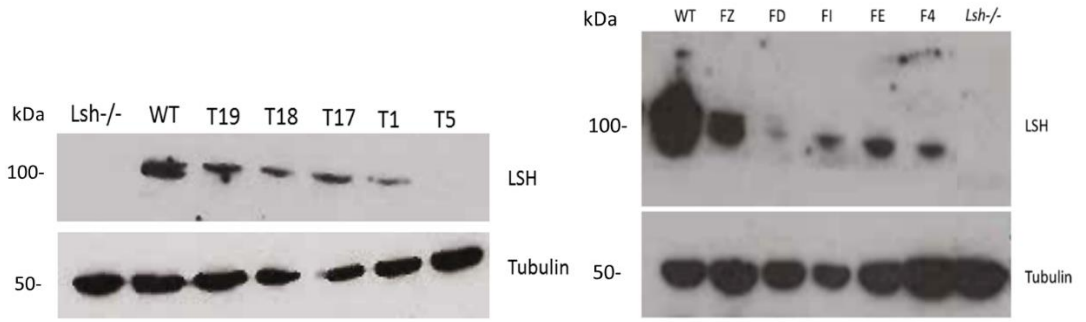


Figure 3.7. Generation of *Lsh* rescue and mutant rescue cell lines. A&B. Schematic of the *PiggyBac* transposon rescue vectors used to transfect *Lsh*$^{-/-}$ ES cells. The rescue vector (A) contains the *Lsh* gene driven by an EF1A promoter and a mCherry/ puromycin resistance dual reporter gene to allow selection of stable expression clones (highlighted in yellow). The mutant rescue vector (B) contains the *Lsh* gene containing a missense mutation in the conserved helicase domain driven by an EF1A promoter and a hygromycin resistance gene to allow selection of stable expression clones (highlighted in yellow). *PiggyBac* transposon terminal repeats (highlighted in pink) permit transposition of the construct into the host genome. The region highlighted in grey contains a pUC origin of replication and an ampicillin resistance gene allowing replication of the plasmid and selection in E.coli. **C&D.** Western blot for LSH protein in rescue cell clone extracts (C) and mutant rescue cell clone extracts (D). Protein extracts from WT and *Lsh*$^{-/-}$ ES cells were used as positive and negative controls. β-TUBULIN antibody was used as a loading control.

3.2.4.2 Neurodifferentiation of rescue cell lines is akin to WT

The *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines were exposed to the same neural differentiation protocol as previously described. Consistently, across 3 different differentiation experiments, morphology changes were more in keeping with the WT than the *Lsh*$^{-/-}$ cell line as there was a lesser abundance of mature neurons at days 7 and 10 (Figure 3.8). This was confirmed by analysing the same lineage markers as previously by qRT-PCR (Figure 3.9). This indicates that both the WT and the mutant version of *Lsh* were able to reverse the *Lsh*$^{-/-}$ cells greater propensity for neural differentiation.

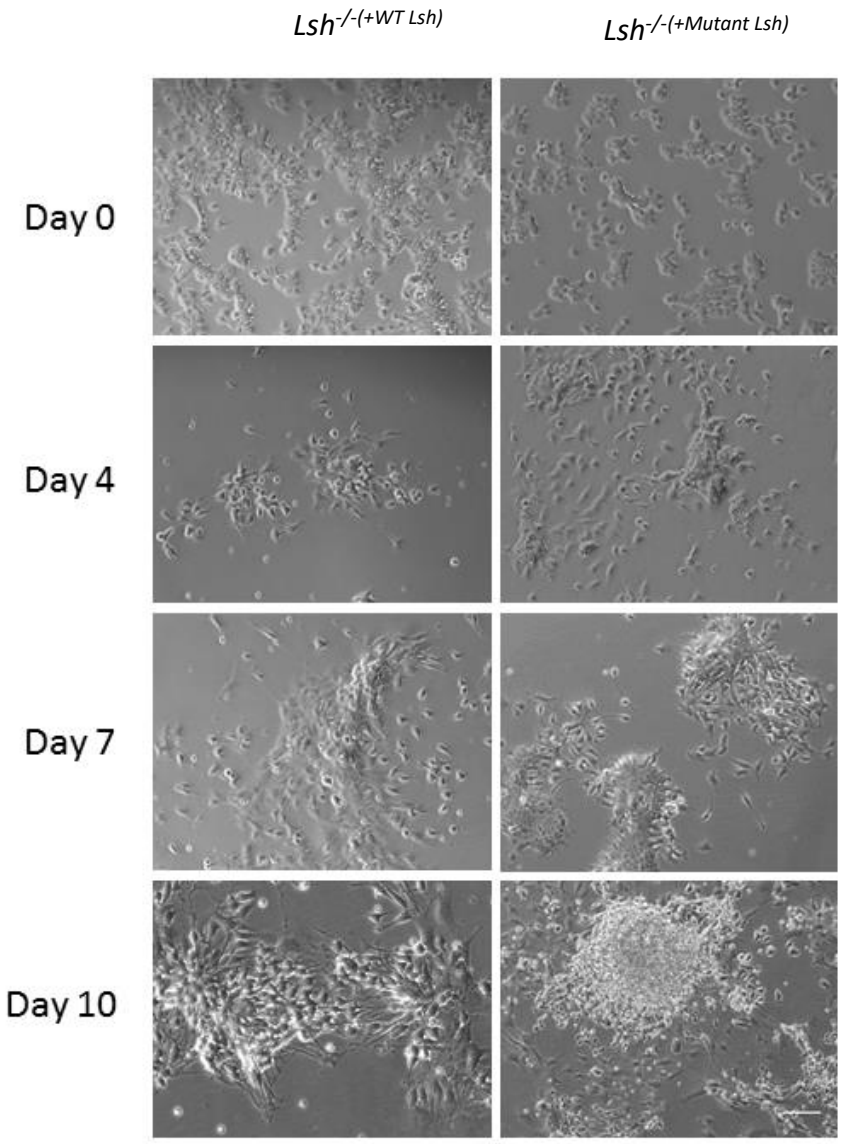


Figure 3.8. Neural differentiation of rescue cell lines. A. Images of *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines at Day 0 in ES serum media and at days 4, 7 and 10 of neurodifferentiation. Scale bars represent 100 µm.

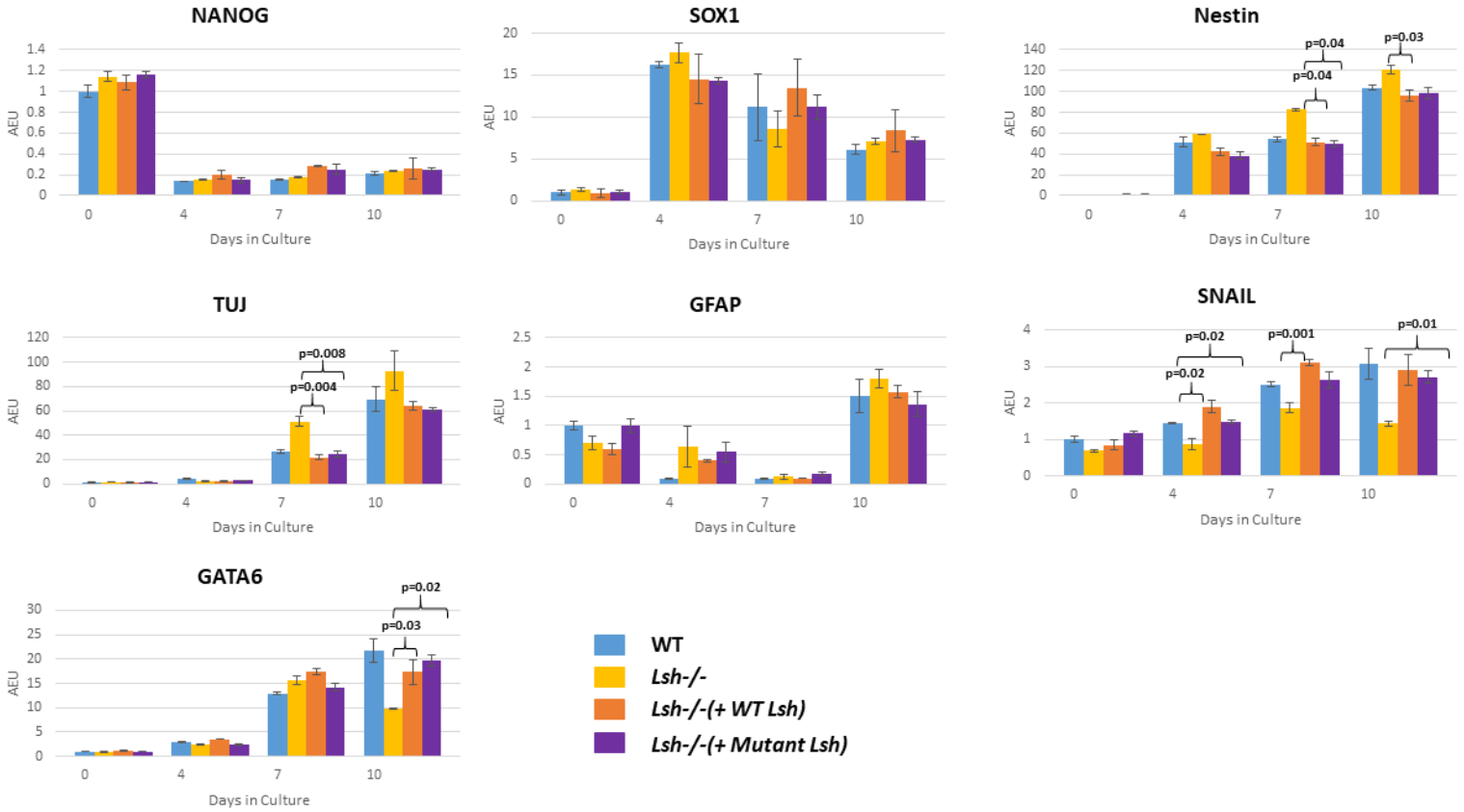


Figure 3.9. Lineage marker expression in rescue cell lines. Quantification of expression of lineage marker genes by qRT-PCR. Expression is normalised to *Gapdh* expression and then standardized to WT ES expression. Error bars represent +/- SEM of 3 biological replicates. P-values indicate level of significance calculated by two-tailed student t-test. Only significant results from comparisons between *Lsh*$^{-/-}$, *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines are shown. AEU=arbitrary expression units.

3.2.4.3 The DNA methylation defect is rescued by re-expression of wild type *Lsh*

The key question that remains to be answered is can the methylation defect seen in *Lsh*$^{-/-}$ NPCs be rescued by the reintroduction of *Lsh*? For this analysis, I assessed DNA methylation at major satellite regions in WT, *Lsh*$^{-/-}$, *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines using bisulphite sequencing to allow quantification of methylation levels at these sites.

Sodium bisulphite treatment of DNA deaminates unmodified cytosine bases to uracil which can then be identified as thymine residues upon sequencing. At the same time methylated cytosines are protected from deamination and hence are read as cytosine residues upon sequencing. Therefore one can generate DNA methylation profiles at single nucleotide resolution. Bisulphite treated DNA from ES and NP cells from all 4 cell lines was amplified using major satellite primers and subjected to Sanger sequencing.

This analysis revealed no gross difference in CpG methylation at major satellites in WT, *Lsh*$^{-/-}$, *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines when maintained as ES cells at 84.4%, 83.3%, 90.2% and 88.9% respectively. WT cells maintained this level of methylation upon differentiation to NPCs remaining at 88.9% methylated. In contrast, *Lsh*$^{-/-}$ cells were unable to maintain their methylation with the number of methylated CpGs reducing to 52.9% following differentiation. This failure to maintain methylation upon differentiation was rescued by the reintroduction of wild type *Lsh,* as the *Lsh*$^{-/-(+WT\ Lsh)}$ rescue cell line maintained methylation levels of 91.1%. The mutated version of *Lsh,* however, was unable to rescue this methylation defect with the *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell line losing methylation upon differentiation such that NPCs had 35.9% of sampled CpGs methylated (Figure 3.10).

This confirms the requirement for *Lsh* for appropriate DNA methylation upon neural differentiation at major satellites.

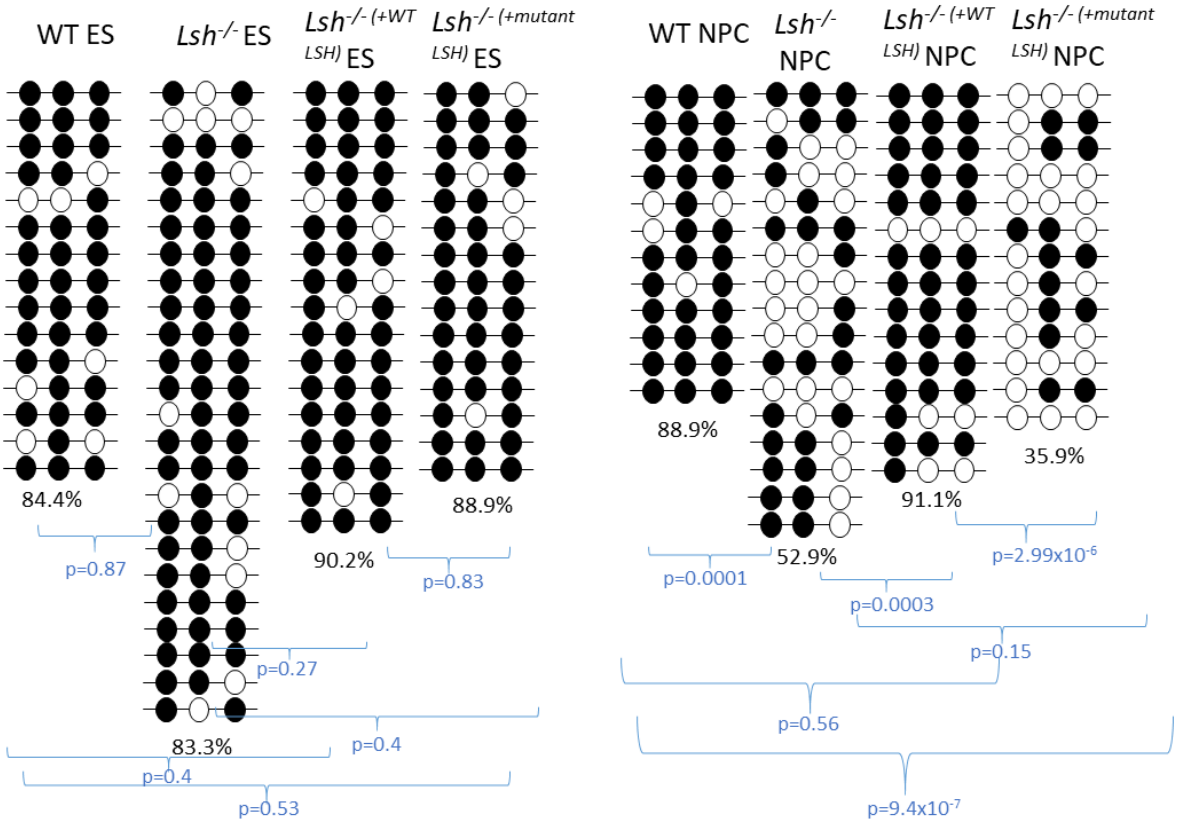


Figure 3.10. DNA methylation at major satellites quantified by bisulphite sequencing. Each circle represents a CpG with black indicating methylated and white indicating unmethylated sites. The percentage of methylated CpGs is shown under each plot. P values display significance using the binomial test in R. Primers were designed against mouse major satellite EF028077 chr9: 3,017,919-3,024,581.

3.3 Discussion

In this chapter, I have described the use of a cell culture system whereby the role of *Lsh* in early neurogenesis could be examined. The N2B27 differentiation protocol by Pollard *et al* was adapted for use as it relies on an autocrine induction mechanism. The monolayer nature of the system also allows easy visualisation of differentiation as it occurs. In my hands, this method was successful in generating neurons as evidenced by typical morphological changes and neural lineage marker expression. Magnetic activated cell sorting allowed for selection of neural progenitor cells within heterogeneous cultures.

The work carried out using this culture system had 2 key aims; to assess if ES cells lacking *Lsh* could develop normally down neural lineage, and to determine if *Lsh* is required for DNA methylation during this process.

3.3.1 *Lsh*$^{-/-}$ cells have a greater propensity to differentiate down neural lineage

During the differentiation protocol I consistently observed a greater abundance of mature neurons, easily identifiable by their long projecting axons, at an earlier stage of differentiation in the *Lsh*$^{-/-}$ cells compared to wild type cultures. This was confirmed by a greater expression of the post-mitotic neural marker *Tuj1,* demonstrated by immunocytochemistry and qRT-PCR. This could be due to a preferential or premature differentiation down neural lineage. The higher expression of both early and later neural lineage markers at day 10 of differentiation in the *Lsh*$^{-/-}$ cells combined with a lower expression of non-neural lineage markers would lend support to a preferential neural differentiation. Further support for this comes from FACs experiments that demonstrated a greater proportion of NPCs in day 10 *Lsh*$^{-/-}$ cultures.

Intriguingly, there are conflicting reports on the importance of *Lsh* in cellular differentiation. *Lsh* has been proposed to play a role in stem cell gene repression via promoter methylation during differentiation, as knock down of *Lsh* by siRNAs prevented complete silencing of stem cell genes and maintained stem cell characteristics *in vitro* (Xi et al., 2009). This was not the case in my cultures where *Nanog* was effectively down-regulated in *Lsh*$^{-/-}$ cells upon differentiation. More in keeping with my findings, Yu *et al* reported *de novo* formation of enhancer elements at hypomethylated sites in *Lsh*$^{-/-}$ MEFS as evidenced by enrichment of H3K4me1. They found that these regions clustered at neuronal lineage genes and reprogramming of these *Lsh*$^{-/-}$ MEFs to induced pluripotent stem cells was followed by enhanced differentiation down neural lineage (Yu et al., 2014a). In contrast, *Lsh*$^{-/-}$ ES cells derived from knockout blastocysts showed comparable rates of neural differentiation to wild type in a study by Ren *et al* (Ren et al., 2015). One possible reason for these different accounts may be the different differentiation systems used and differing timing of assement

of differentiation. For example Ren *et al* used a differentiation protocol of embryoid body formation followed by retinoic acid treatment therefore is not directly comparable to my work. My system does have an advantage over the others in that it relies on an autocrine induction mechanism, therefore may better reflect the effects of lack of *Lsh* in driving differentiation. It is, of course, artificial as differentiation *in vivo* is also driven by extracellular signals.

Also it should always be borne in mind that different cell lines will show inherent differences in their ability to differentiate. For this reason it was important to repeat my experiment after re-expression of *Lsh* in the knockout cell line. The fact that this rescued the differentiation differences lends support to the argument that it is the lack of *Lsh* itself that led to the differences seen in neural differentiation.

In the next chapter, I will explore possible reasons for this greater propensity for neural differentiation. One possibility we considered was that there may be differences in the *Lsh*$^{-/-}$ methylome at developmental genes. Intriguingly, however, the re-expression of *Lsh* with a mutation in a conserved helicase domain, which did not rescue the methylation defect, at least at major satellites, did rescue the differentiation phenotype. This suggests that the effects of *Lsh* upon differentiation may be independent of DNA methylation.

3.3.2 *Lsh* is required for DNA methylation during neurogenesis

At the beginning of this project, little was known about the developmental time point at which *Lsh* acted to ensure appropriate levels of DNA methylation. In this chapter, I have demonstrated a key role for *Lsh* in DNA methylation during differentiation of pluripotent embryonic stem cells to neural progenitor cells as demonstrated by LC-MS, restriction enzyme digest and bisulphite sequencing. During my time carrying out research on this project a paper was published corroborating this finding (Ren et al., 2015).

Of particular interest is the fact that global DNA methylation levels as seen by LC-MS were lower in *Lsh*$^{-/-}$ NPCs than *Lsh*$^{-/-}$ ES cells. This suggests a role for *Lsh* in maintenance methylation as opposed to its widely reported role in *de novo* DNA methylation. An alternative explanation could be that there are further demethylation/remethylation events at this developmental timepoint.

Having confirmed the requirement for *Lsh* in DNA methylation during early neurogenesis I was interested in assessing exactly where in the genome this methylation defect occurred as the majority of previous studies have focused only on specific loci. Furthermore, I was intrigued to examine whether this methylation defect had any transcriptional consequences. These points will be addressed in chapter 4.

Chapter 4. Mapping the *Lsh*$^{-/-}$ methylome, transcriptome and H3K27me3 deposition upon neurodifferentiation

4.1 Introduction

In chapter 3, I demonstrated the requirement for *Lsh* for normal DNA methylation during neurogenesis by assessing global DNA methylation levels. In this chapter, I intend to further scrutinise the genomic sites at which *Lsh* contributes to methylation at this developmental stage, by mapping the genome-wide DNA methylome. There are very few previous studies that examine the genome wide methylome in *Lsh*$^{-/-}$ models and none using neural cells.

I will then go on to answer a key question; is this methylation defect linked with transcriptional consequences? There are several reports of up-regulated genes in *Lsh*$^{-/-}$ models, which authors have accredited to hypomethylation of single copy gene promoters and suggested that *Lsh* plays a role in targeting DNA methylation to these sites (Fan et al., 2005b; Myant et al., 2011; Xi et al., 2009; Xi et al., 2007).

Despite the established role for *Lsh* in ensuring appropriate satellite repeat DNA methylation, less attention has been paid to the transcriptional consequences of repeat hypomethylation in *Lsh*$^{-/-}$ models. Although previously considered as “junk DNA” there is a growing recognition of the importance of these regions in genome integrity and regulation (Garcia-Perez et al., 2016). I therefore intend to explore the transcriptional consequences of repeat hypomethylation further at this early stage of neurogenesis.

4.2 Results

4.2.1 DNA methylome mapping by ERRBS

In order to establish the regions of the genome at which *Lsh* contributes to DNA methylation during neurogenesis, genome-wide 5-mC profiles generated by Enhanced Reduced Representation Bisulphite Sequencing (ERRBS) were analysed. “Reduced representation” refers to the prior digestion of the DNA with a MspI restriction enzyme, followed by size fractionation to select for CG-rich DNA fragments (Meissner et al., 2005). This method therefore provides coverage of important features of the genome such as promoters, whilst reducing the total volume of sequencing required by other methods such as whole genome bisulphite sequencing. The obvious disadvantage to this reduced coverage is the lack of information it provides about methylation at non-CGIs. Enhanced RRBS increases the coverage of RRBS by also including larger DNA fragments post MspI digest, which have a lower CG content, and using a whole genome alignment approach rather than aligning to the Msp1 fragmented genome. This is reported to result in a 75% increase in the coverage of CpG sites, providing information not only about promoter regions but also about introns, exons and CpG shores (the regions immediately flanking CpG islands) (Akalin et al., 2012; Garrett-Bakelman et al., 2015).

DNA from one replicate each of *Lsh*$^{-/-}$ and WT ESC and NPC samples, were sent to the Epigenomics Core Facility of Weill Cornell Medicine who performed ERRBS and processed the raw data. Data was then mapped to the mm9 build of the mouse genome and analysed by Dr Donncha Dunican.

4.2.1.1 *Lsh* is required for DNA methylation of repetitive elements during neurogenesis

As experiments in chapter 3, and previously published reports, had established a role for *Lsh* in DNA methylation of satellite sequences, I initially scrutinized the ERRBS dataset for methylation over different repeat subclasses. Only CpGs mapping to long repeats, considered as full length repeats, were included in this robust analysis (defined as >6kb for LTRs and LINE1s and >1kb for satellites). This was done to prevent the overwhelming inclusion of reads from multitudes of short retrotransposon remnants with unknown biological relevance. In contrast, full-length repeats would be predicted to have a greater consequence upon local chromatin environment and possibly retain the potential to retrotranspose.

To calculate repeat methylation the average methylation of individual CpGs, with a minimal depth of 5 reads, mapping to long repeats were plotted in violin plots (Figure 4.1).

Surprisingly, LINE1s and LTRs were marginally, but significantly, hypomethylated in *Lsh*$^{-/-}$ ES cells compared to WT at 40% vs 46% ($p<2.2x10^{-16}$) and 79% vs 86% ($p=2.4x10^{-11}$)

respectively. In satellite repeats, significant hypomethylation was not seen with *Lsh*$^{-/-}$ ES cells showing 73% compared to 79% (p=0.87) methylation in WT. It should be noted, however, that that the number of CpGs in the satellite repeat dataset is far less than the number of LINE1s or LTRs (139 vs 33127 and 2870 respectively) which will clearly have an effect on significance testing.

Upon differentiation to NPCs, LINE1s, LTRs and satellite repeats all became grossly hypomethylated in *Lsh*$^{-/-}$ cells compared to WT at 50% vs 77.4% , 47.1% vs 86.9% and 46.2% vs 75% respectively. Concentrating on the change in methylation upon differentiation, one can see that in wild type cells, LTR and satellite sequences maintained their methylation level whereas LINE1s gained methylation. In the knockout cell line, LINE1s also gained methylation upon differentiation but failed to do so to the same extent as wild type cells. This is in keeping with the reported role *Lsh* plays in facilitating *de novo* methylation. However, the knockout cells actually lose methylation at LTR and satellite sequences upon differentiation lending further support for a role for *Lsh* in maintenance of DNA methylation. However, one cannot exclude the occurrence of demethylation / remethylation events at these sites during ESC to NPC transition. Repeating these experiments at multiple time points during this developmental window would be required to confirm this.

The role *Lsh* plays in DNA methylation is clearly context dependent. Despite its high expression in ES cells it appears largely dispensable for DNA methylation until these cells differentiate to neural progenitors. Although there is a small degree of hypomethylation in ES cells, the levels are not consistent with *Lsh* playing a major role in maintenance of methylation at this stage.

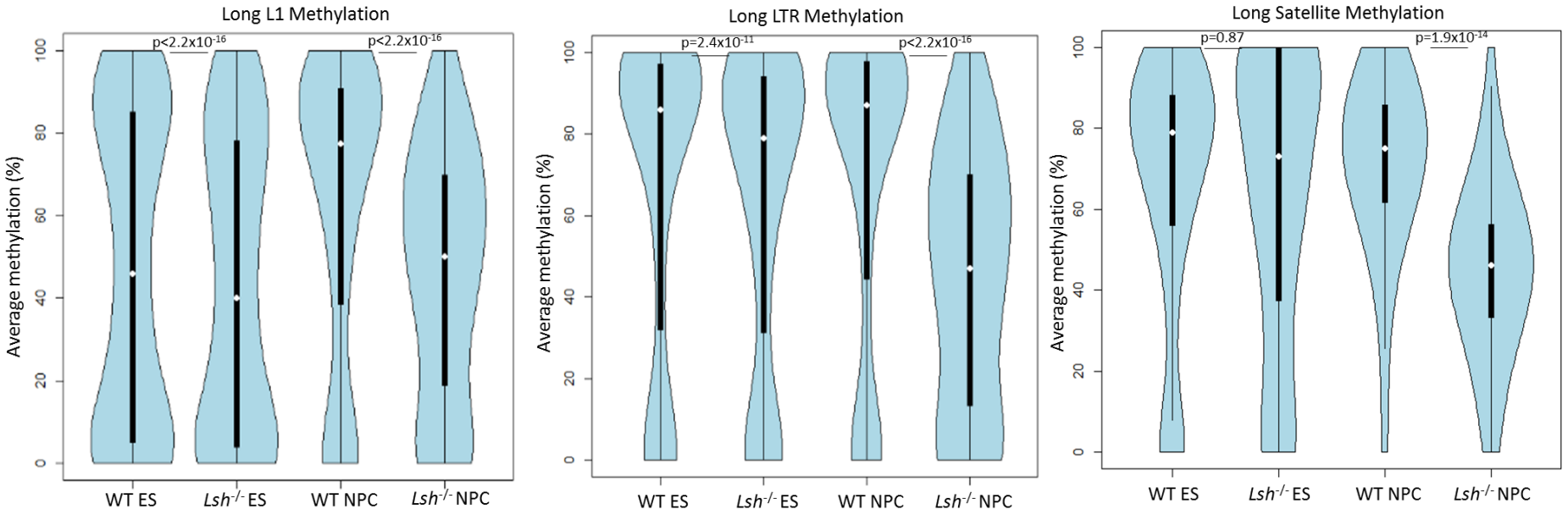


Figure 4.1. Repeats are hypomethylated in *Lsh*$^{-/-}$ *cells*. Violin plots demonstrate the distribution frequency with which individual CpGs within a family of full length repeats are methylated in WT and *Lsh*$^{-/-}$ ESCs and NPCs. White spots indicate median values and black bars represent the interquartile range. P values show results of significance testing by Wilcoxon rank sum test.

4.2.1.2 Gene promoters associated with immune function are hypomethylated in *Lsh*$^{-/-}$ NPCs

As mentioned earlier, previous reports exist of *Lsh* playing a role in DNA methylation at promoters of single copy genes, including those of stem cell genes (Fan et al., 2005b; Xi et al., 2009). I therefore explored promoter methylation in the ERRBS data set. Promoter methylation was calculated as the average methylation of all CpGs spanning the +/- 2kb region across the TSS.

In order to define promoters that lost or gained methylation, a threshold change of 20% absolute methylation difference was set. Although arbitrary, this threshold was chosen to make results comparable to other work published in this area (Velasco et al., 2018b). This revealed very few differentially methylated promoters between WT and *Lsh*$^{-/-}$ ES cells, with 41 showing increased and 66 decreased methylation compared to WT. However, following differentiation 395 promoters showed a more than 20% reduction in methylation in *Lsh*$^{-/-}$ NPCs and only 39 demonstrated a gain in methylation compared to WT (Figure 4.2).

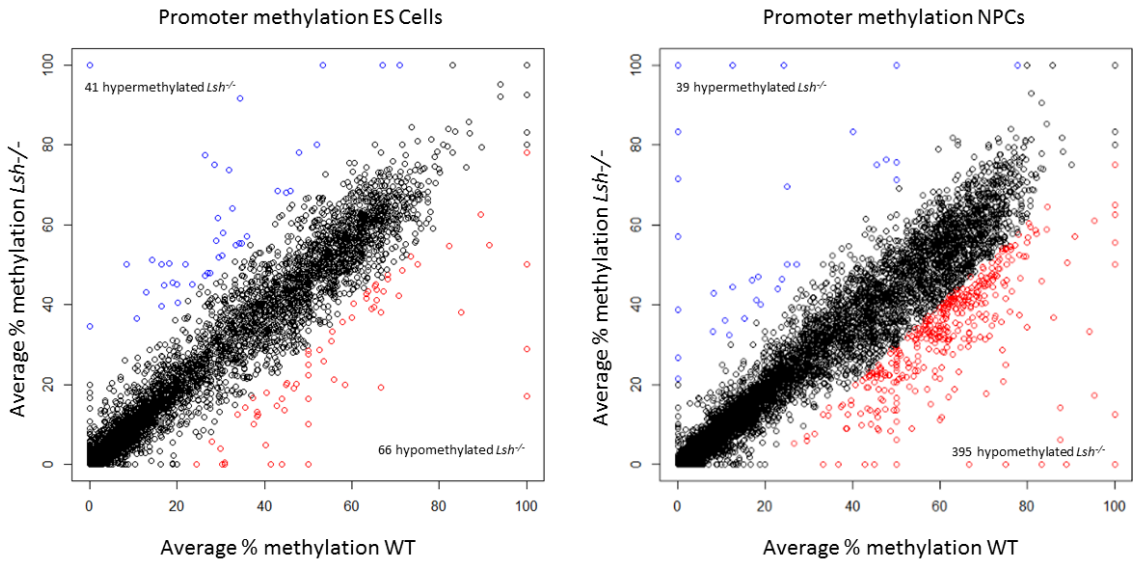
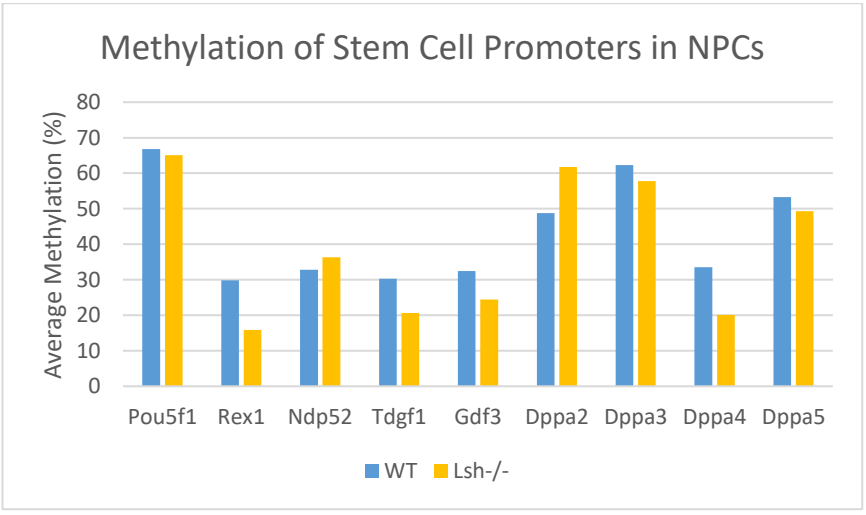


Figure 4.2. Promoter methylation in *Lsh*$^{-/-}$ cells. The scatter plots display the average methylation across promoters (defined as +/- 2kb from TSS) in WT and *Lsh*$^{-/-}$ cells. Differentially methylated promoters (defined as >20% methylation difference) are highlighted in red for those that were hypomethylated and blue for those that were hypermethylated in *Lsh*$^{-/-}$ cells.

In chapter 3, I demonstrated that ES cells lacking *Lsh* displayed a greater propensity for neural differentiation compared to wild type. I was keen to see if this could be explained by differences in promoter methylation at key pluripotency or lineage commitment genes. Contrary to previous reports (Xi et al., 2009), *Lsh* does not appear to be required for de novo methylation of promoters of stem cell specific genes (as defined in the Xi et al paper) upon differentiation as *Lsh*$^{-/-}$ NPCs displayed similar levels of methylation to WT (Figure 4.3 A). This is in keeping with my finding that there was no inhibition of differentiation capability of the *Lsh*$^{-/-}$ ES cells in chapter 3.

Gene ontology analysis of the hypomethylated gene promoters in NPCs did not reveal an enrichment for neural differentiation pathways, demonstrating that the preferential neural differentiation seen in the *Lsh*$^{-/-}$ cell line was not due to a widespread hypomethylation of neural lineage genes. Instead ontology analysis revealed an enrichment of terms to do with inflammatory and innate immune responses (Figure 4.3 B). In saying this, 2 genes known to play a role in differentiation were found to be hypomethylated in both *Lsh*$^{-/-}$ ES and NP cells; namely *Nefm* (a neurofilament protein involved in nervous system development) used as an NPC marker and *Ttll4* (involved in polyglutamylation of tubulin proteins) which plays a role in early embryonic development and possibly in neurite outgrowth (Song and Brady, 2014; Ye et al., 2018). Therefore one cannot completely exclude this as a contributing mechanism.

A



B

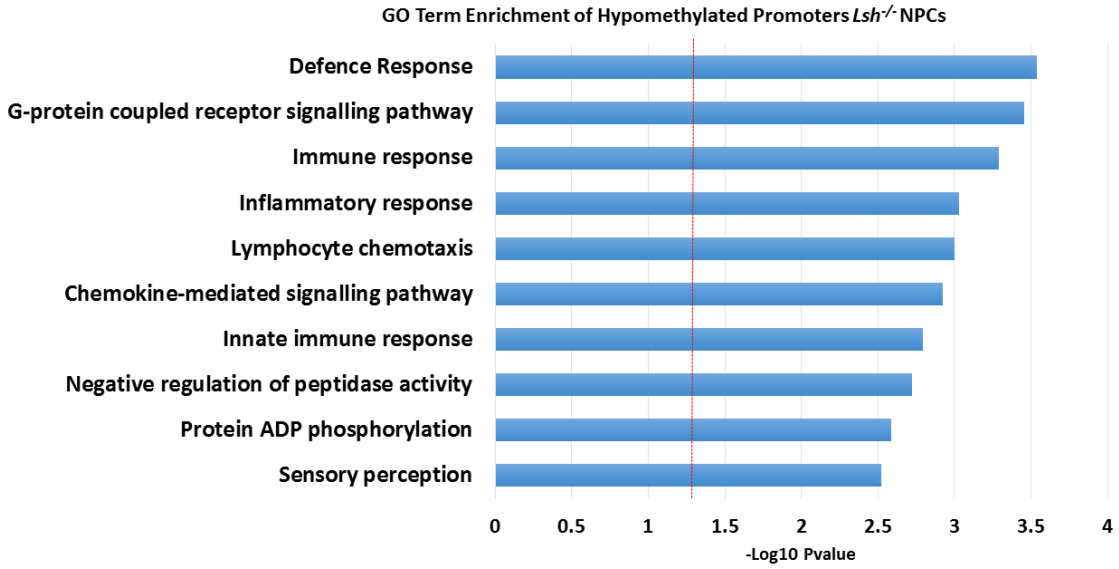


Figure 4.3. Innate immune genes display hypomethylation in *Lsh*$^{-/-}$ NPCs. A. Boxplot displays average methylation across stem cell promoters in NPCs from ERRBS data. **B**. Gene ontology analysis of hypomethylated promoters in *Lsh*$^{-/-}$ NPCs. Table shows the top ten most significant biological process terms as found using David V6.8. The red dashed line demarcates significance value of p=0.05.

4.2.1.3 L1 repeats are enriched at hypomethylated promoters in *Lsh*$^{-/-}$ NPCs

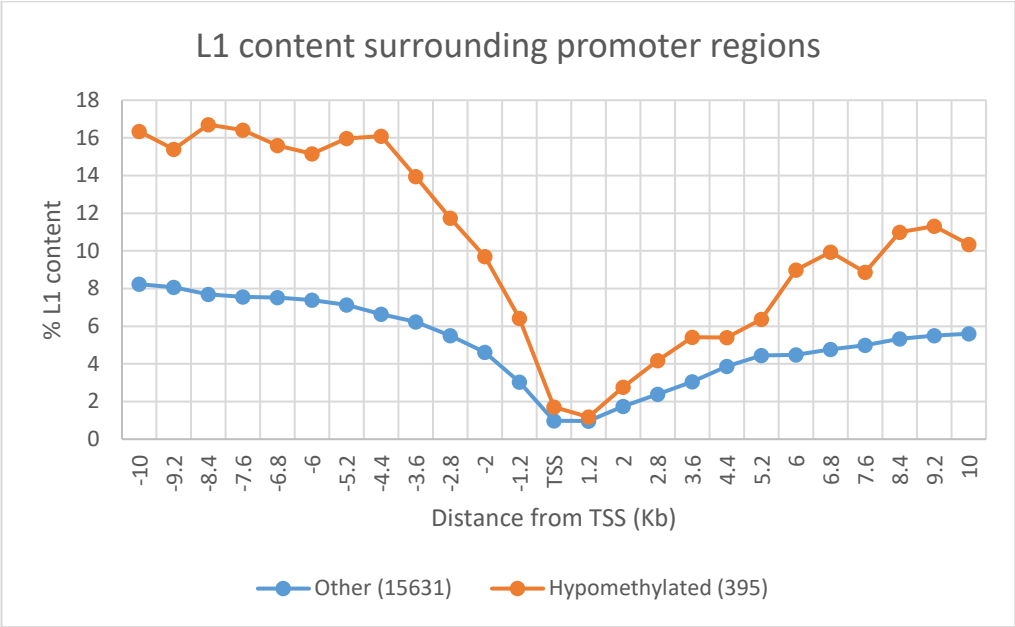
The ERRBS data has clearly shown that *Lsh* is required for repeat methylation upon neurodifferentiation, but also that a number of gene promoters, particularly those of immune response genes, are hypomethylated in the absence of *Lsh*. A possible explanation for the methylation defect seen at these promoters is that they are in a region of the genome rich in repeat content and are therefore hypomethylated due to a position effect rather than by direct targeting by *Lsh*.

To investigate this I analysed the repeat content of the genome surrounding these hypomethylated promoters. This was carried out by dividing the genome into 800bp windows, 10kb up and downstream of TSSs of genes that were hypomethylated in *Lsh*$^{-/-}$ NPCs, and those genes with no change in promoter methylation. These files were then intersected with repeat masker files from the UCSC genome browser. Percentage repeat content was analysed using the IRanges package in R and displayed as the percentage of base pairs within an 800bp window that are part of an annotated repeat.

Analysis was carried out for all annotated L1, LTR and satellite repeats regardless of length. Data from satellite repeats is not shown as the number of these in the vicinity of gene promoters was too small to extract meaningful data.

This analysis demonstrated that the 20kb stretch of genome surrounding TSSs of hypomethylated promoters was enriched for L1 elements, when compared to regions surrounding promoters that displayed no difference in promoter methylation between *Lsh*$^{-/-}$ and WT NPCs. This was also true, although to a lesser extent, for LTR content at these regions (Figure 4.4).

This could indicate that, in contrast to what previous authors have reported when examining single copy loci, *Lsh* does not target specific promoters for methylation, but rather these promoters become hypomethylated in the absence of *Lsh* due to loss of methylation at nearby repeat targets.



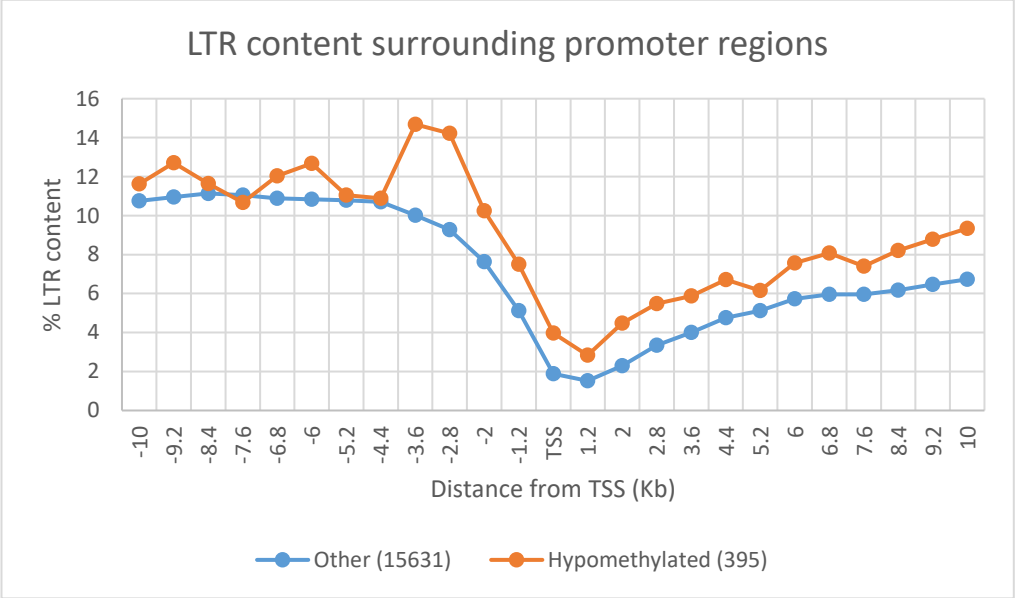


Figure 4.4. Repeat content of genome surrounding hypomethylated promoters in *Lsh*$^{-/-}$ NPCs. This graph divides the genome surrounding the TSS of genes with hypomethylated promoters in *Lsh*$^{-/-}$ NPCs (orange) and genes with unchanged promoter methylation status when compared to WT (blue), into 800bp bins 10kb up and downstream of the TSS. Each bin is represented by a point on the graph. The average percentage of base pairs within each 800bp bin that form part of an annotated L1 or LTR repeat, regardless of length, is shown on the y axis. The number of genes in each group is indicated in parenthesis.

4.2.2 RNA-Seq analysis of the transcriptome

A key question that arises is: does this wide scale reduction in methylation have any effects upon transcription? In order to assess the effects, if any, of lack of *Lsh* on the transcriptome, I isolated polyadenylated mRNA and prepared libraries from WT and *Lsh*$^{-/-}$ ES and NP cells and sent samples to Edinburgh Genomics for Illumina sequencing. Each sample had 3 biological replicates. Raw data was processed by Dr Donncha Dunican.

4.2.2.1 Repeats are not up-regulated in *Lsh*$^{-/-}$ NPCs

Having established gross hypomethylation of repeat sequences, it seemed logical to analyse their transcription. As previously stated, transposable elements are usually strictly repressed in somatic cells, although there is evidence for more activity in the brain compared to other somatic cells for reasons as yet unclear (Goodier, 2016). We analysed repeat expression within the RNA-Seq dataset for both reads aligning to full length (long) LINE1, LTR or satellite sequence and those reads aligning to non-full length (short) repeats, as defined in section 4.2.1.1. In many cases a "short" repeat is actually likely to form part of a long, not fully annotated, repeat therefore their inclusion in the analysis is highly relevant. Furthermore aberrant transcription from retrotransposon remnants retains the potential to influence expression of host genes. A repeat was considered differentially expressed if there was a greater than 4 fold change in expression and FDR <0.05.

In ES cells, full length repeats showed no differential expression with only 2 L1s showing some degree of up-regulation. When the analysis was extended to short repeats, again there was no evidence for repeat de-repression (Figure 4.5). Despite hypomethylation occurring at these sites upon neurodifferentiation, there was also no evidence of up-regulation in repeat transcription in *Lsh*$^{-/-}$ NPCs (Figure 4.6). qRT-PCR confirmed that there was no significant change in repeat expression in *Lsh*$^{-/-}$ cells (Figure 4.7).

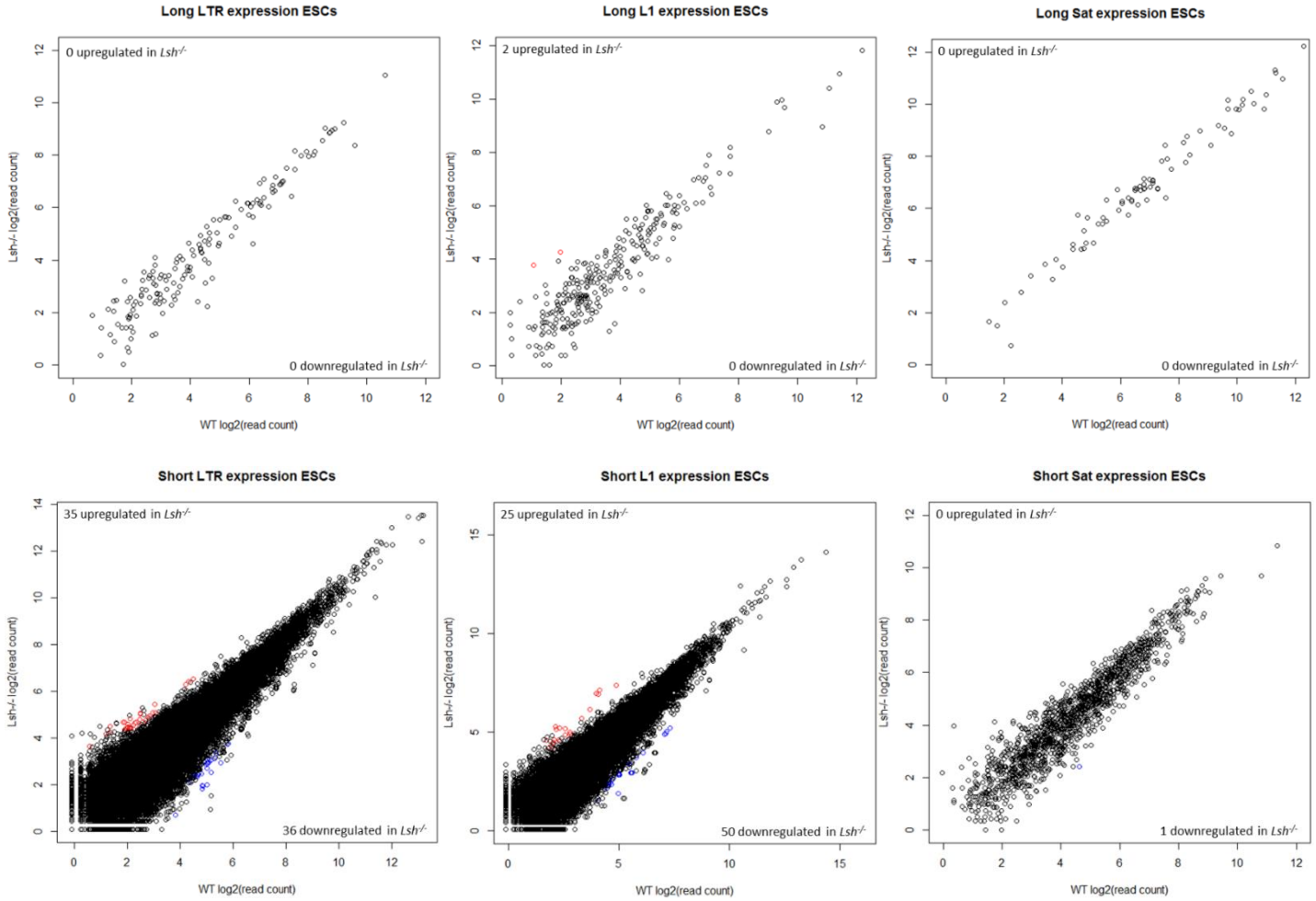


Figure 4.5. Repeat Expression in ESCs. Scatter plots display read counts of LINE1 (L1), long terminal repeat (LTR) and satellite (Sat) sequences from RNA-Seq analysis from 3 biological replicates of WT and *Lsh*$^{-/-}$ ES cells. Repeats defined as differentially expressed in *Lsh*$^{-/-}$ cells (>4 fold change and FDR<0.05) are highlighted in red if up-regulated and blue if down-regulated. Data is shown for reads mapping to full length repeats (long) and non-full length repeats (short).

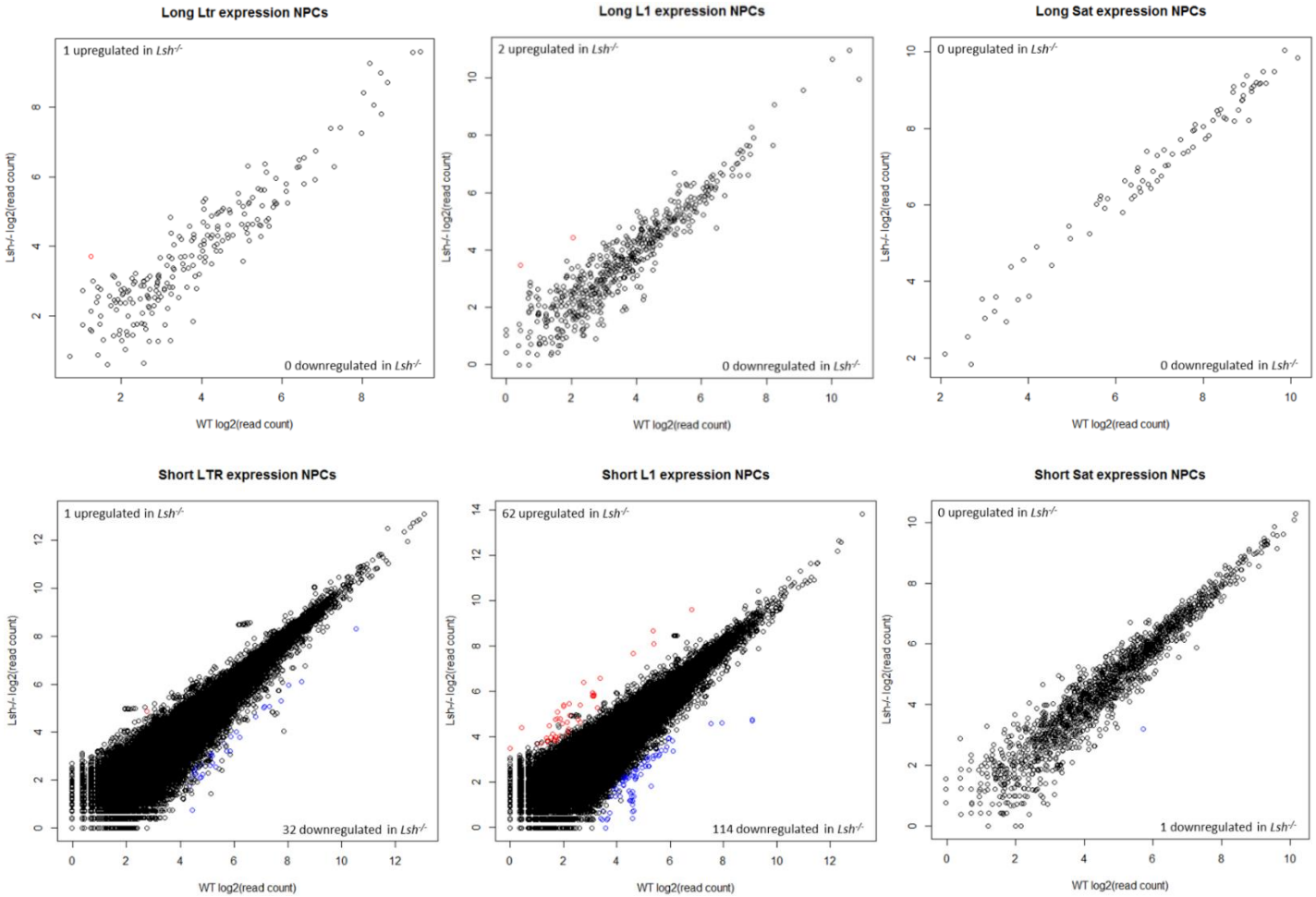


Figure 4.6. Repeat Expression in NPCs. Scatter plots display read counts of LINE1 (L1), long terminal repeat (LTR) and Satellite (Sat) sequences from RNA-Seq analysis from 3 biological replicates of WT and *Lsh*$^{-/-}$ NP cells. Repeats defined as differentially expressed in *Lsh*$^{-/-}$ cells (>4 fold change and FDR<0.05) are highlighted in red if up-regulated and blue if down-regulated compared to WT. Data is shown for reads mapping to full length repeats (long) and non-full length repeats (short).

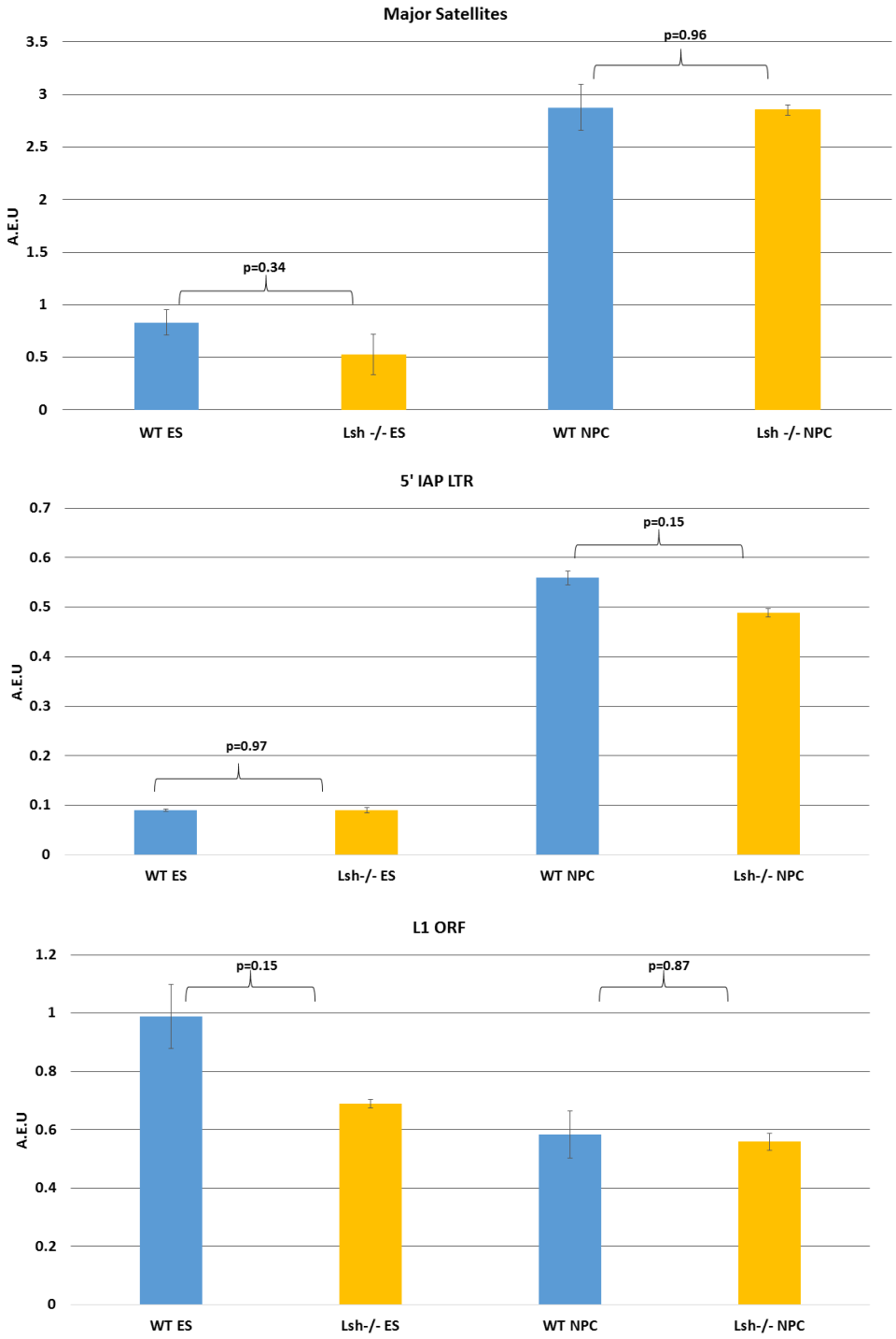


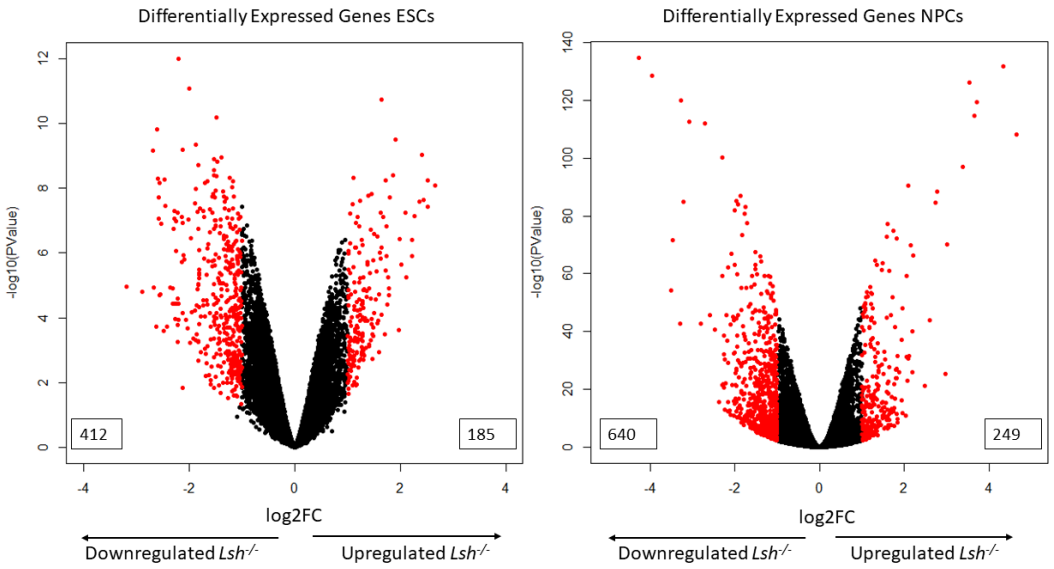
Figure 4.7. qRT-PCR of repeat expression. Bar charts display results of qRT-PCR for major satellites, LINE1s and LTRs. The error bars represent the SEM from 3 biological replicates. The Y axis displays arbitrary expression units (AEU) and results are normalised to *Gapdh*. No significant difference was seen between WT and *Lsh*$^{-/-}$ cells as calculated by student t test.

4.2.2.2 Differential gene expression in the absence of *Lsh*

I then went on to examine differential gene expression in the RNA-Seq data (defined as >2 fold change in expression and P value < 0.05 across 3 biological replicates). This revealed a total of 597 differentially expressed genes in *Lsh*$^{-/-}$ ES cells with 412 genes showing down-regulation and 185 up-regulation. The reason for this differential expression in ES cells when *Lsh* does not appear to be required for DNA methylation is unclear, and points to additional roles for *Lsh* at this pluripotent stage when it is certainly highly expressed. The number of differentially expressed genes increased following differentiation to 640 genes showing down-regulation and 249 up-regulation (Figure 4.8A). The fact that there is a skew to down-regulation as opposed to up appears to be contradictory to the global hypomethylation seen in *Lsh*$^{-/-}$ NPCs. This fact in itself highlights the growing recognition that, for the majority of genes, transcription is not governed by DNA methylation. The relationship between DNA methylation and gene transcription will be explored further in later sections. There was a small but significant degree of overlap between those genes differentially expressed in *Lsh*$^{-/-}$ ES cells and *Lsh*$^{-/-}$ NPCs. 33 genes that were down-regulated in *Lsh*$^{-/-}$ ESCs were also down-regulated in *Lsh*$^{-/-}$ NPCs (more than would be expected by chance, p=0.002 by hypergeometric testing) and 11 genes that were up-regulated in *Lsh*$^{-/-}$ ESCs were also up-regulated in *Lsh*$^{-/-}$ NPCs (p=0.0005) (Figure 4.8B). The fact that the majority of genes did not overlap demonstrates an important role for *Lsh* in regulating gene transcription during the differentiation process.

In order to determine if any particular class of genes were misexpressed, gene ontology analysis was carried out using David V6.8 utilising the biological processes category (Figure 4.9). Analysis of down-regulated genes in *Lsh*$^{-/-}$ ESCs revealed an enrichment of terms related to transcriptional processes whereas up-regulated genes were mainly to do with differentiation processes suggesting that in the absence of *Lsh,* ES cells may be more "primed" for differentiation. In neural cells developmental terms also featured strongly in the ontology analysis. This highlights a role for *Lsh* in ensuring appropriate timing of developmental gene expression and may explain the differentiation differences seen in chapter 3.

A



B

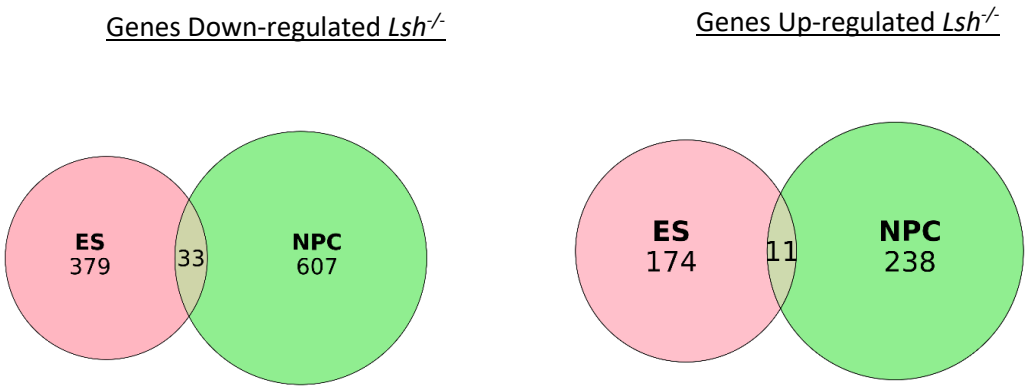


Figure 4.8. Differential gene expression in *Lsh*$^{-/-}$ cells. A. Volcano plots comparing transcriptomes of WT and *Lsh*$^{-/-}$ ES and NP cells respectively. Differentially expressed genes (>2 fold difference in expression and $p < 0.05$) are highlighted in red. Total numbers of differentially expressed genes are provided in the boxes. **B.** Venn diagrams display the overlap of differentially expressed genes in *Lsh*$^{-/-}$ ES and NP cells with total gene numbers displayed (overlapping gene lists are displayed in Appendix List A1&2).

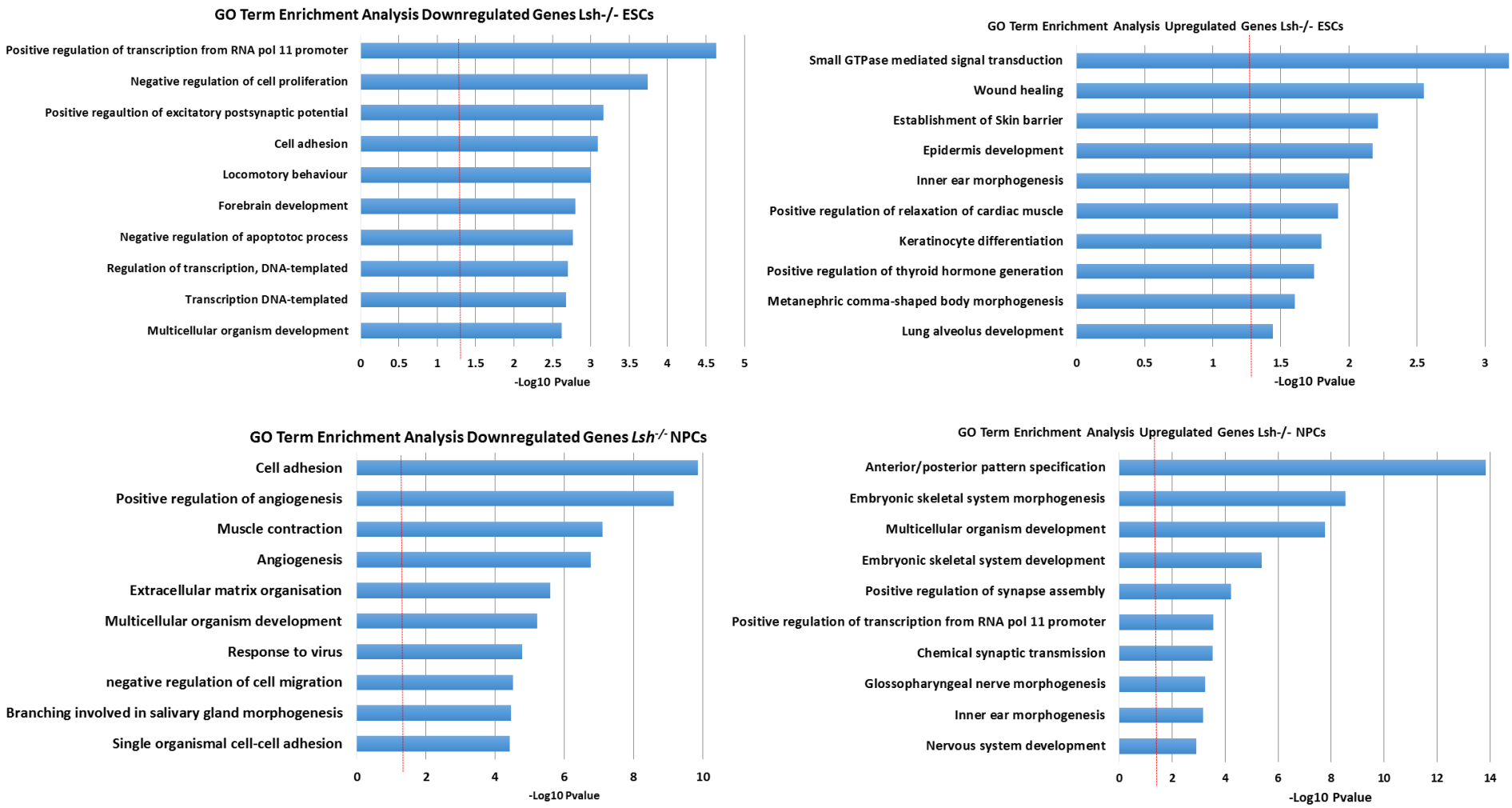


Figure 4.9. Gene ontology of differentially expressed genes. Charts show the top ten most significantly enriched gene ontology terms of differentially expressed genes in *Lsh*$^{-/-}$ ES and NP cells as analysed by the biological process direct tool of David V6.8. Red line indicates p value =0.05.

4.2.2.3 Differential methylation at promoter regions does not account for transcriptional changes

As demonstrated earlier, the global hypomethylation that occurs upon differentiation in *Lsh*$^{-/-}$ cells did not appear to be associated with a widespread up-regulation in gene expression. However, the question remains as to whether the hypomethylation seen over individual promoters in the absence of *Lsh* results in increased transcription from these genes. To examine this, I integrated my ERRBS and RNA-Seq datasets. This revealed that of the 118 genes that overlapped between these datasets from ES cells, only 2 of the genes that were defined as transcriptionally up-regulated were also defined as having promoter hypomethylation; namely *Nefm* and *Ttll4* (previously discussed). Similarly, only 5 of the overlapping 184 genes defined as up-regulated in NPCs were also defined as having hypomethylated promoters; namely *Nefm, Ttll4, Cldn15* (a component of tight junction complexes*), Fv1 (*a retroviral restriction gene*)* and *Gm5124* (a pseudogene) (Figure 4.10).

One can conclude from this, that the vast majority of genes up-regulated in the absence of *Lsh* are not so due to promoter hypomethylation. I chose a rather robust definition of a reduction in 20% across the whole promoter, reasoning that anything less than this would be unlikely to alter local chromatin conformation and access for the transcriptional machinery. However it is unknown to what degree methylation has to change to alter transcription and this would depend on chromatin context. Indeed methylation loss at even a single CpG could affect transcription if this was at for example a transcription factor binding site (Yin et al., 2017).

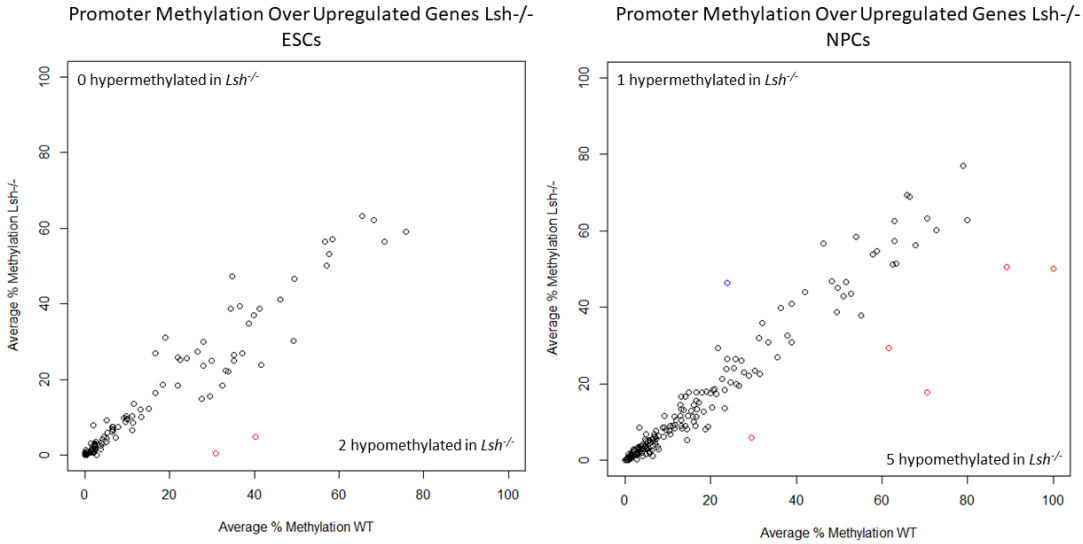


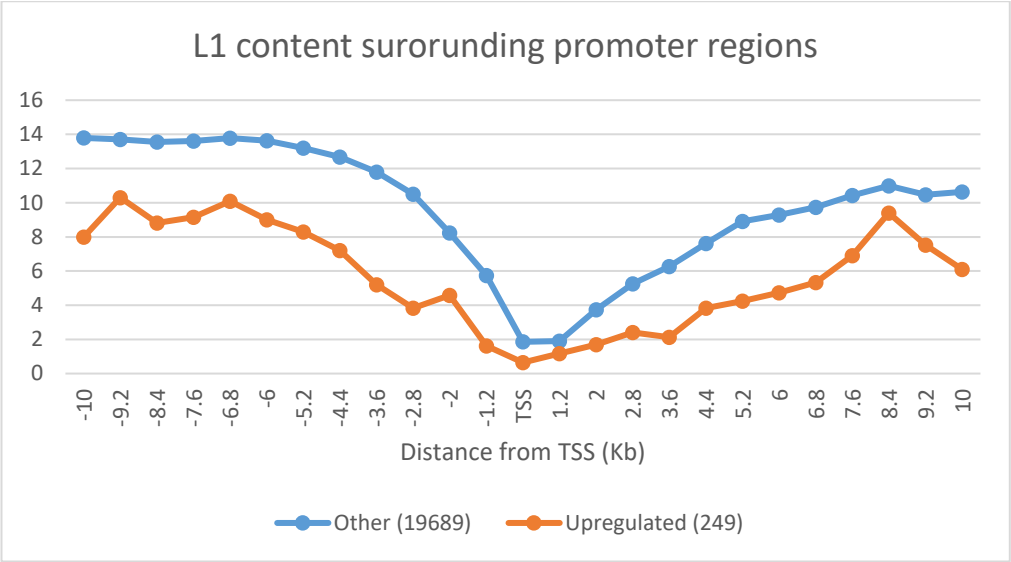
Figure 4.10. Lack of correlation between transcriptional upregulation and promoter hypomethylation. The scatter plots display the average promoter methylation of genes up-regulated in *Lsh*$^{-/-}$ ES and NP cells. Those genes with differentially methylated promoters (defined as a change of >20%) are shown as red circles for hypomethylated and blue for hypermethylated in *Lsh*$^{-/-}$ cells.

4.2.2.4 Repeats are depleted near genes up-regulated in *Lsh*$^{-/-}$ NPCs

Another possible explanation for differential gene expression in the knockout cell line would be hypomethylation of repeats creating a more transcriptionally permissive environment for nearby genes. I therefore interrogated the RNA-Seq data to analyse if those genes that were up-regulated in *Lsh*$^{-/-}$ NPCs tended to be in a more repeat rich region of the genome than non-differentially expressed genes. In order to do this, I utilised the same methods as previously described in section 4.2.1.3. Analysis was carried out for all annotated L1 and LTR repeats regardless of length. Again data from satellite repeats is not shown as the number of these in the vicinity of gene promoters was too small to extract meaningful data.

Results are shown as the percentage of base pairs which form part of an annotated repeat in 800bp windows up to 10kb up and downstream of the TSS (Figure 4.11). This revealed that, as a whole, up-regulated genes actually had fewer L1 or LTR repeats in close proximity to their TSS than genes that were not differentially expressed.

The data presented here and in section 4.2.2.3, suggests that DNA methylation is not the primary factor regulating the differential gene expression.



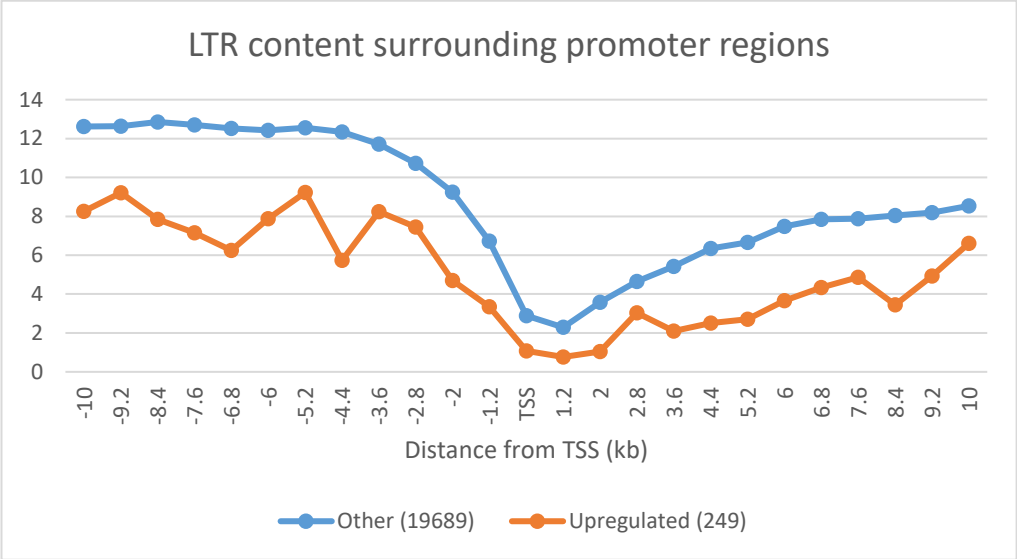


Figure 4.11. Repeat content of genome proximal to up-regulated genes in *Lsh*$^{-/-}$ NPCs. This Graph represents the repeat content of the genome proximal to the TSS of genes up-regulated in *Lsh*$^{-/-}$ NPCs (orange line) compared to those genes not differentially expressed (blue line). The region of genome analysed is 10kb up and downstream of the TSS divided into 800bp bins. Each bin is represented by a point on the graph. The average percentage of base pairs within each 800bp bin that form part of an annotated L1 or LTR repeat, regardless of length, is shown on the y axis. The number of genes in each group is indicated in parenthesis.

4.2.2.5 *Hox* genes are misexpressed in the absence of *Lsh*

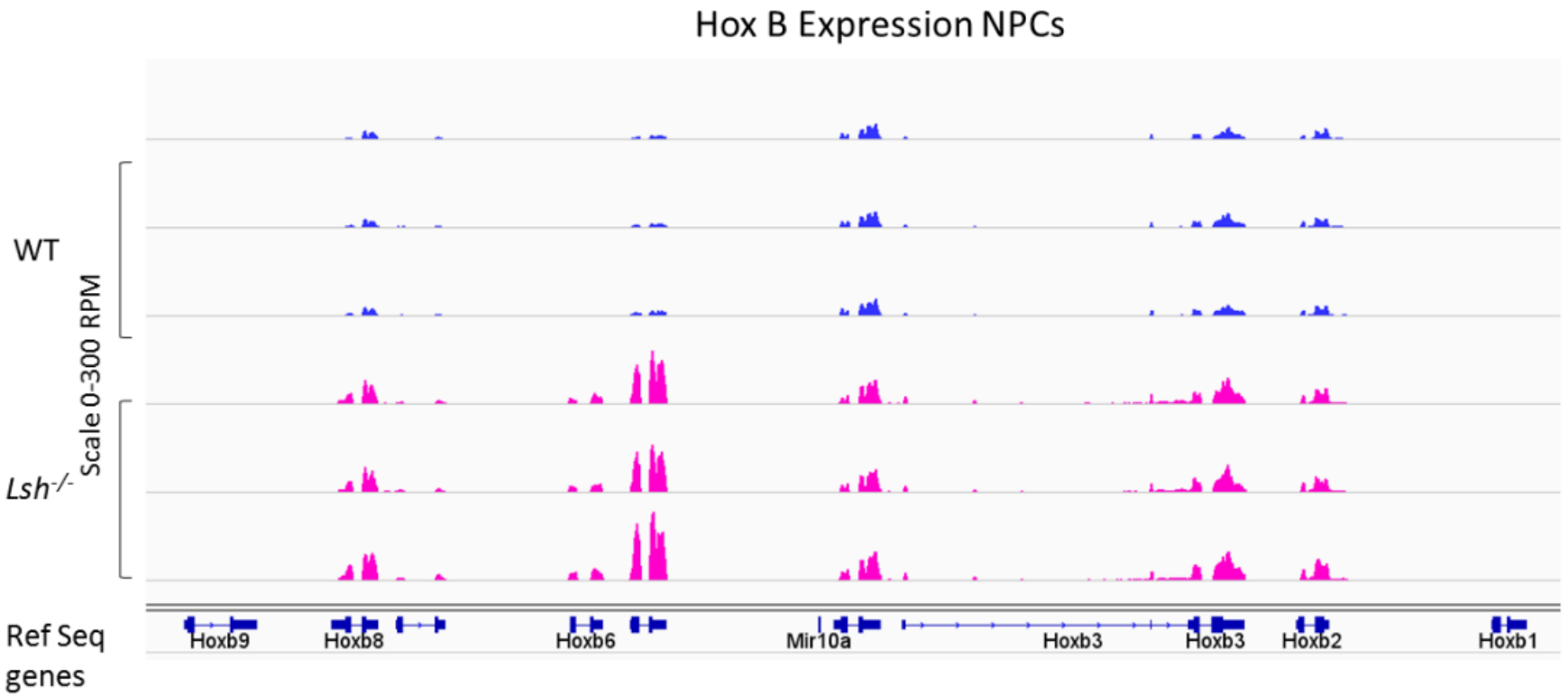
Of particular note was the up-regulation of developmentally regulated *Hox* genes in *Lsh*$^{-/-}$ NPCs (Figure 4.12). These genes encode homeobox transcription factors which play key roles in development, as their downstream targets direct cellular and positional identity of differentiating cells along the anterior-posterior axis of the embryo (Mallo and Alonso, 2013). *Hox* genes are organised into clusters of which there are 4 in the mouse termed A-D, and expression is tightly regulated in a spatiotemporal manner during development, in part by epigenetic changes (Soshnikova and Duboule, 2009).

Their expression has been found to be up-regulated in other models lacking *Lsh* and it has been suggested that this up-regulation is due, at least in part, to hypomethylation at *Hox* gene promoters (Xi et al., 2007). In my ERRBS analysis however, *Hox* gene promoters in *Lsh*$^{-/-}$ cells were not found to be hypomethylated. Although DNA methylation may play a role in silencing some *Hox* genes during development, the key epigenetic regulatory system at these gene clusters is the Polycomb repressive complex, which mediates deposition of the repressive mark, H3K27me3 (Barber and Rastegar, 2010).

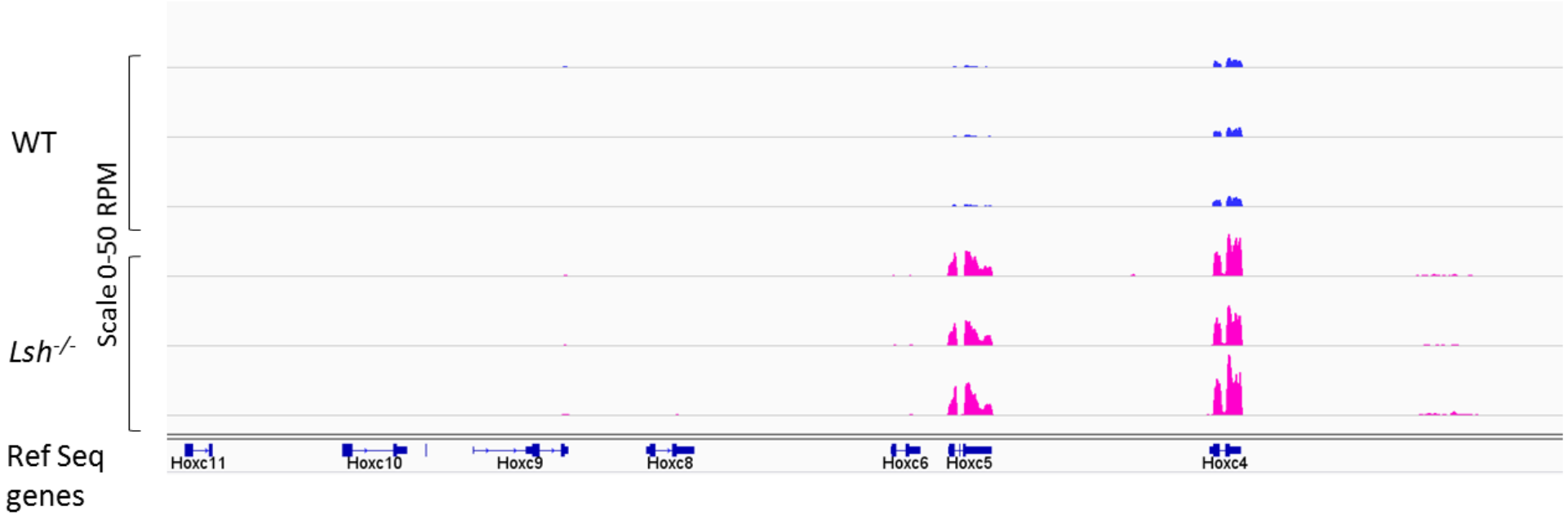
I was therefore interested to investigate if there were any abnormalities in the Polycomb system in the *Lsh*$^{-/-}$ NPCs which could account for the up-regulation of *Hox* genes. Also, I hypothesised that deficits in this system could explain the differential expression of other genes that are not accounted for by changes in DNA methylation.

Figure 4.12


Hox A Expression NPCs
WT
Lsh-/-
Scale 0-40 RPM
Ref Seq genes
Hoxa1
Hoxa2
Hoxa3
2700086A05Rik
Hoxa7
Hoxa9


Hox B Expression NPCs
WT
Lsh-/-
Scale 0-300 RPM
Ref Seq genes
Hoxb9
Hoxb8
Hoxb6
Mir10a
Hoxb3
Hoxb3
Hoxb2
Hoxb1

Hox C Expression NPCs



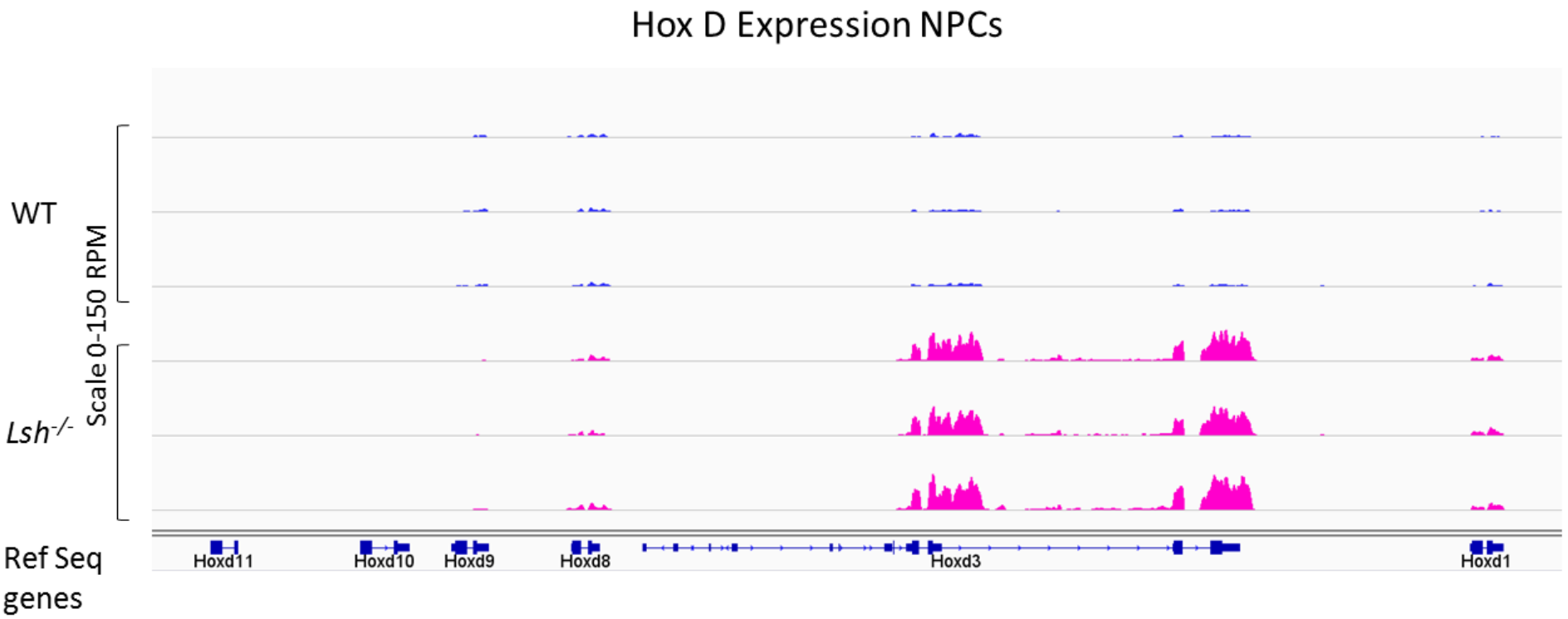


Figure 4.12. Hox Genes are misexpressed in *Lsh*$^{-/-}$ NPCs. IGV genome browser shots displaying RNA-Seq expression data across Hox clusters A-D from 3 biological replicates of WT and *Lsh*$^{-/-}$ NPCs

4.2.3 Analysis of H3K27me3, a Polycomb repressive mark, by ChIP-seq

So far, I have focused only on DNA methylation during development. However, the finding of up-regulated expression of developmental genes in my RNA-Seq analysis prompted me to investigate the distribution of Polycomb repression in my model system. Polycomb group proteins form two major repressive complexes, PRC1 and PRC2, and provide a further level of epigenetic control of transcription via post translational modification of histone N-terminal tails. The repressive H3 modification of H3K27me3 is deposited by the histone methyltransferase *Ezh2/Ezh1*, a member of the PRC2 complex.

Similarly to DNA methylation, chromatin modifications are also highly dynamic during development, with modifications associated with active chromatin reducing globally concomitant with an increase in repressive histone marks as development proceeds (Meshorer and Misteli, 2006). Developmentally important genes, such as *Hox* genes, display "bivalent" chromatin marks in ES cells, i.e. both active H3K4me3 and inactive H3K27me3. This maintains them in a non-expressed, but poised state ready for transcriptional activation upon differentiation, thereby maintaining their pluripotent status (Bernstein et al., 2006).

As cells differentiate, a proportion of these bivalently marked genes lose either their repressive or active mark, dependent upon the specific lineage. Some that are silenced may also be associated with DNA methylation. Some cells will lose the repressive mark and become active during differentiation only to regain the mark at later stages of differentiation. Almost all *Hox* genes are bivalently marked in ES cells. Upon in vitro differentiation to NPCs, *Hox* A1/4/5 are known to lose H3K27me3 to become actively marked with the other *Hox* genes either remaining bivalent or losing their active mark to become marked by H3K27me3 only. In mature brain tissue all are marked by H3K27me3 (Meissner et al., 2008a). *Hox* genes are however expressed in the brainstem and spinal cord and have central roles in neuronal subtype specification within regional domains (Philippidou and Dasen, 2013).

In order to investigate the distribution of H3K27me3 I performed ChIP-seq. In this approach histone proteins are chemically cross-linked with formaldehyde to bound DNA. The cells are then lysed, chromatin fragmented by sonication and an antibody against H3K27me3 is then used to pull down the regions of bound DNA. Following preparation of libraries, one sample from each genotype in each cellular context (i.e. ESC or NPC) was sent to Edinburgh Genomics for next generation Illumina sequencing and bioinformatic analysis was carried out by Dr Donncha Dunican.

I carried out quality control checks in my ChIP samples. Firstly, gel electrophoresis of sonicated chromatin was carried out for each sample to ensure appropriate and equal fragment size. Sonication was optimised to result in chromatin fragments of 100-300bp (Figure 4.13 A). This fragment length is a compromise, being long enough to allow unique

mapping but short enough to improve specificity and give tighter ChIP-seq peaks. In order to assess the specificity of the H3K27me3 antibody, qRT-PCR was performed on samples incubated with the antibody and those incubated with a non-specific antibody to IgG. H3K27me3 is known to be enriched over *Hox* genes in ES cells therefore *Hox C9* was selected as a positive control region. The active housekeeping gene β-actin, which is devoid of H3K27me3 in ES cells, was selected as a negative control region. Results confirmed enrichment over positive control regions with low background from IgG pull down (Figure 4.13 B).

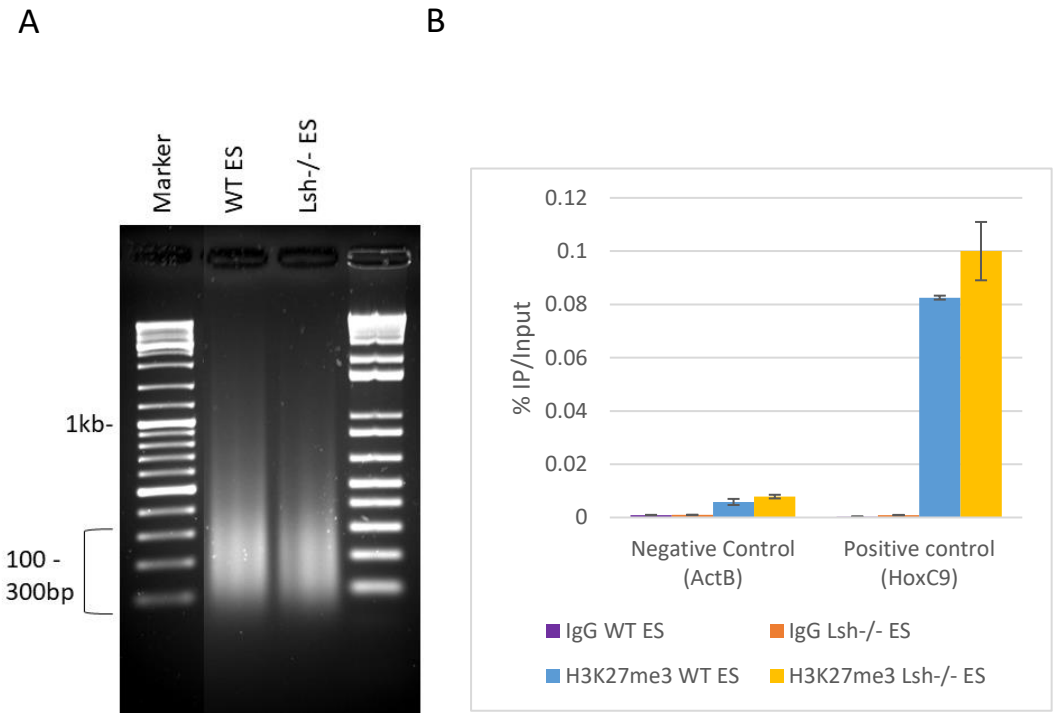


Figure 4.13. ChIP quality control. A. A representative agarose gel following sonication of cross-linked chromatin demonstrating fragments of 100-300bp. **B.** ChIP-qPCR at positive and negative control regions for H3K27me3 in ES cells. Enrichment is shown as percentage of input pulled down. Error bars represent +/- SEM of three biological replicates.

4.2.3.1 There are widespread losses of the Polycomb repressive mark, H3K27me3, from promoter regions in *Lsh*$^{-/-}$ cells following neurodifferentiation

I previously hypothesised that the increased expression of *Hox* genes in NPCs, in the absence of *Lsh,* could be due to a loss of Polycomb repressive marks from these sites. I therefore looked at H3K27me3 across promoter regions of *Hox* genes from the ChIP-Seq dataset, defined as +/- 2kb from TSS, and indeed found a small reduction in the mark across all 4 *Hox* clusters (Figure 4.14 A).

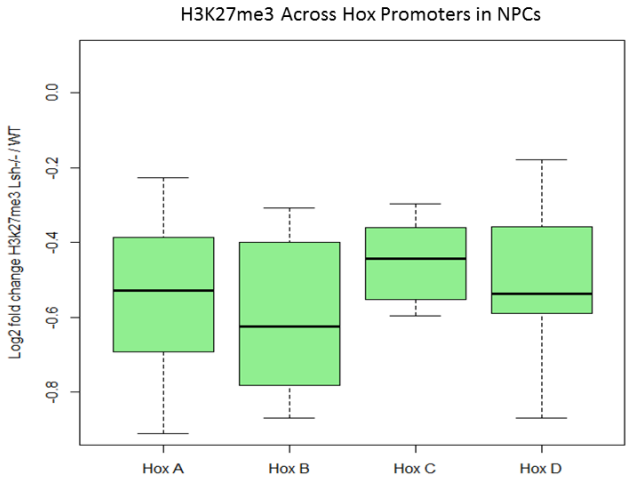
Rather than looking at candidate genes only, I then went on to examine H3K27me3 across all promoter regions in the dataset. H3K27me3 across a promoter was defined as the average number of reads from 1kb sliding windows (sliding by 200bp) over the +/- 2kb region from the TSS. In order to define promoters that lost or gained H3K27me3 in *Lsh*$^{-/-}$ compared to WT, a threshold of 2 fold change was set excluding all promoters with less than 10 reads per million. This method revealed that very few promoters in *Lsh*$^{-/-}$ ES cells had altered Polycomb occupancy, with only 9 showing a gain in H3K27me3 and 13 a reduction. Following differentiation however, *Lsh*$^{-/-}$ cells demonstrated 5 promoters with increased H3K27me3 and a substantial 248 with reduced mark demonstrating that Polycomb regulation is altered in *Lsh*$^{-/-}$ NPCs (Figure 4.14 B).

As can be seen in Figure 4.14 B, not only do these 248 promoters have a >2 fold reduction in H3K27me3 compared to WT NPCs, but also there appears to be a more global reduction in H3K27me3 across promoters.

Interestingly, gene ontology terms for those promoters with >2 fold reduction in H3K27me3, were to do with neural differentiation (Figure 4.14 C). It is therefore tempting to hypothesise that as cells differentiate, the absence of *Lsh* results in a redistribution of Polycomb away from traditional target sites, including neurodevelopmental genes, resulting in their premature activation and so accounting for the preferential neural differentiation seen in chapter 3.

Figure 4.14

A



B

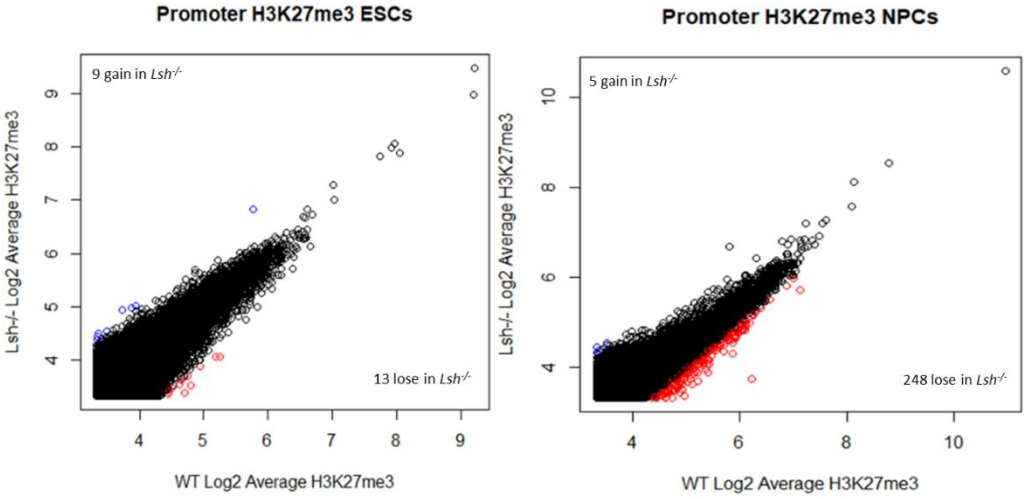


Figure 4.14. Polycomb regulation is altered in *Lsh*$^{-/-}$ NPCs. **A.** Boxplot demonstrates the log2 fold change in reads from H3K27me3 ChIP across *Hox* clusters in *Lsh*$^{-/-}$ NPCs compared to WT. A negative value demonstrates reduction of the mark in *Lsh*$^{-/-}$ cells. **B.** Scatter plots display the log2 value of the average number of reads from H3K27me3 ChIP across promoter regions of genes in *Lsh*$^{-/-}$ and WT ES and NP cells. The number of reads across a promoter was taken as the average number of reads from 1kb sliding windows (sliding by 200bp) +/- 2kb from TSS. Promoters that had a greater than 2 fold change in *Lsh*$^{-/-}$ cells compared to WT are shown in blue for those that were enriched in *Lsh*$^{-/-}$,and red for those enriched in WT

C

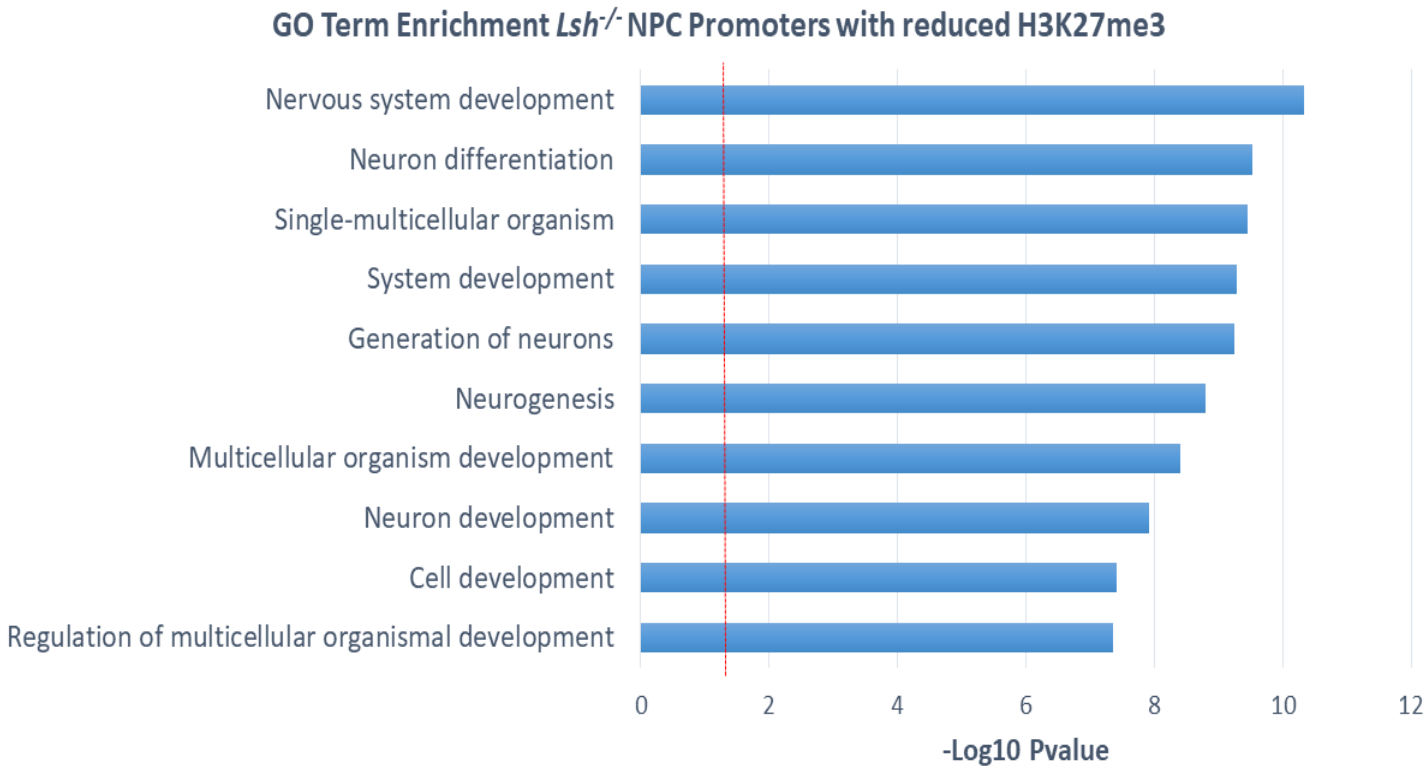


Figure 4.14. Polycomb regulation is altered in *Lsh*$^{-/-}$ NPCs. C. Top ten most significant gene ontology enrichment terms of promoters with >2fold reduction in H3K27me3 in *Lsh*$^{-/-}$ NPCs as analysed by the biological process tool of David V6.8. The red line demarcates p value=0.05.

4.2.3.2 H3K27me3 is reduced over differentially expressed genes

As revealed earlier in this chapter, the changes in gene expression seen in *Lsh*$^{-/-}$ cells were not due to alterations in promoter DNA methylation status. However, having demonstrated disruption of the Polycomb repressive system in the absence of *Lsh,* I was keen to test the hypothesis that alterations in this system may correlate with gene transcription.

In order to test if those genes that were differentially expressed in *Lsh*$^{-/-}$ NPCs were also differentially marked by H3K27me3, I integrated my ChIP-seq and RNA-Seq data sets. This analysis revealed a greater reduction in H3K27me3 across promoters of genes that were differentially expressed compared to those genes which maintained WT expression levels. Interestingly this occurred in both up and down regulated genes (Figure 4.15 A). Furthermore, when looking at WT NPC levels of H3K27me3 across promoters of those genes differentially expressed in *Lsh*$^{-/-}$ NPCs, it was revealed that differentially expressed genes are normally enriched for Polycomb when compared to non-differentially expressed genes (Figure 4.15 B). This confirms that differentially expressed genes in *Lsh*$^{-/-}$ NPCs are Polycomb targets. This implies that the initial hypothesis - that there is a redistribution of Polycomb away from gene promoters resulting in increased expression - although possibly accounting for some differential expression, is too simplistic. This is because the down-regulated genes are also Polycomb targets and lose H3K27me3 in the absence of *Lsh*. It could be that these down-regulated Polycomb targets are so as a consequence of repressive signalling pathways caused by the up-regulation of other Polycomb targets. What can be concluded is that at this early stage of differentiation, *Lsh* plays a key role in H3K27me3 distribution and regulation of Polycomb targeted developmental genes.

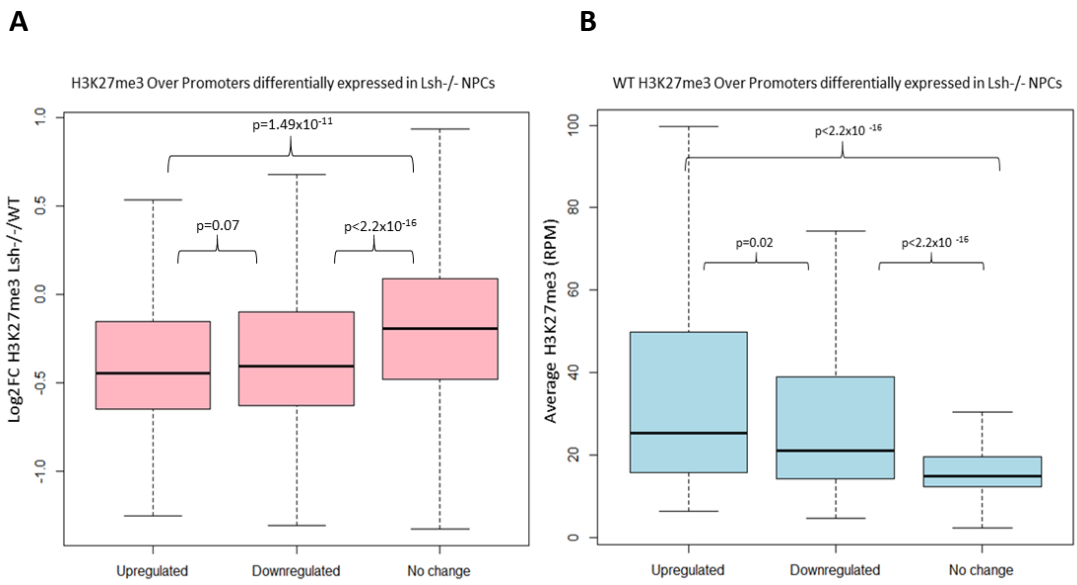


Figure 4.15. H3K27me3 over differentially expressed genes. A. Log2 fold change in H3K27me3 over promoters of genes differentially expressed in *Lsh*$^{-/-}$ vs WT NPCs. Negative values indicate reduced polycomb in *Lsh*$^{-/-}$ cells. **B.** Analysis of WT NPC levels of H3K27me3 across promoters of genes differentially expressed in *Lsh*$^{-/-}$ NPCs reveals that these genes are Polycomb targets. P values indicate significance from wilcoxon rank sum test.

4.2.3.3 H3K27me3 is redistributed to repeats in the absence of *Lsh*

This loss of H3K27me3 from target sites has been demonstrated in other hypomethylated models including *Dnmt1*$^{-/-}$ MEFs (Reddington et al., 2013). This does not appear to be due to any loss of activity of Polycomb as demonstrated by consistent global levels of H3K27me3, but instead appears to be due to a redistribution of the mark. Dunican *et al* demonstrated this in *Lsh*$^{-/-}$ MEFs, where H3K27me3 was found to be increased over repeat sequences (Dunican et al., 2013). As DNA methylation and H3K27me3 are usually mutually exclusive modifications in the genome, rarely being found in the same place, it is hypothesised that loss of methylation causes an increase in potential binding sites for Polycomb thereby "diluting" the mark across the genome.

Having found a loss of H3K27me3 over promoter regions I then examined the mark over full length repeats in order to test this theory of redistribution in my model. This was done by averaging reads across individual full length satellites, LTRs and LINE1s. This revealed no difference in H3K27me3 over repeats between *Lsh*$^{-/-}$ and WT ES cells, however following neurodifferentiation these sequences became enriched for H3K27me3 in *Lsh*$^{-/-}$ cells (Figure 4.16).

This would be in keeping with a loss of methylation at repeats, occurring upon differentiation in the absence of *Lsh*, opening alternative binding sites for Polycomb thereby resulting in its distribution away from target sites to these hypomethylated regions.

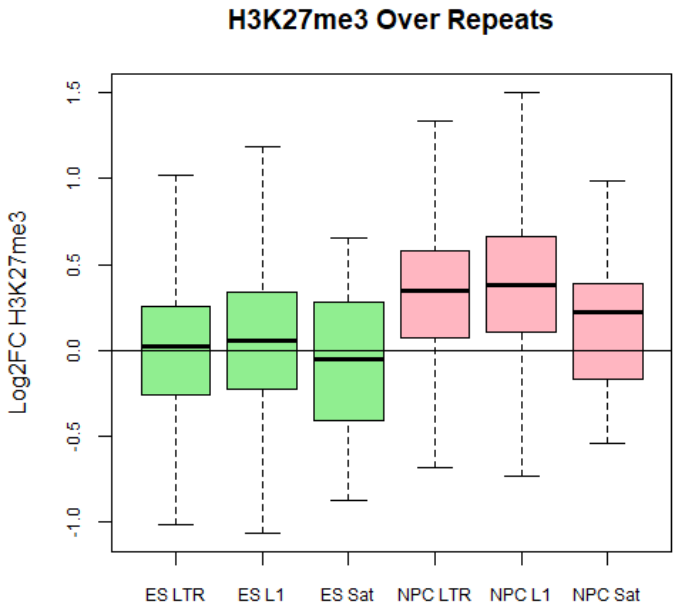


Figure 4.16. Polycomb is redistributed to repeats in *Lsh*$^{-/-}$ NPCs. Boxplot displays the log2 fold change in reads from H3K27me3 ChIP across full length repeats in *Lsh*$^{-/-}$ vs WT ES and NP cells. A positive value indicates an increase in H3K27me3 in *Lsh*$^{-/-}$ cells.

4.2.3.4 Re-expression of wild type and mutated *Lsh* restores *Hox* gene repression

To further test the hypothesis that *Lsh* plays a role in regulating expression of developmental Polycomb target genes, I performed qRT-PCR for *Hox* gene expression in the *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines both when maintained as ES cells and following NPC differentiation.

This revealed that re-expressing wild type *Lsh* in the knockout cells resulted in a rescue of repression of *Hox* genes in NPCs. Interestingly; re-expressing the mutated version of *Lsh* also rescued *Hox* gene repression (Figure 4.17).

This argues against the theory that the up-regulated transcription of developmental genes in *Lsh*$^{-/-}$ NPCs is due to a redistribution of Polycomb complexes away from these sites due to hypomethylation. This is because the *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell line failed to rescue the methylation defect (Section 3.2.4.3).

From this one could hypothesise either, that *Lsh* plays a direct role in targeting Polycomb to developmental genes or it alters transcription at these sites via interaction with other epigenetic pathways independent of its role in DNA methylation.

As these hypotheses are drawn from results of *Hox* gene expression only, in the future they will need to be tested further by ChIP and RNA-Seq in the $Lsh^{-/-(+WT\ Lsh)}$ and $Lsh^{-/-(+Mutant\ Lsh)}$ rescue cell lines. The fact that $Lsh^{-/-(+Mutant\ Lsh)}$ cells behaved similarly to WT on neurodifferentiation would support restoration of developmental gene transcriptional regulation in these cells.

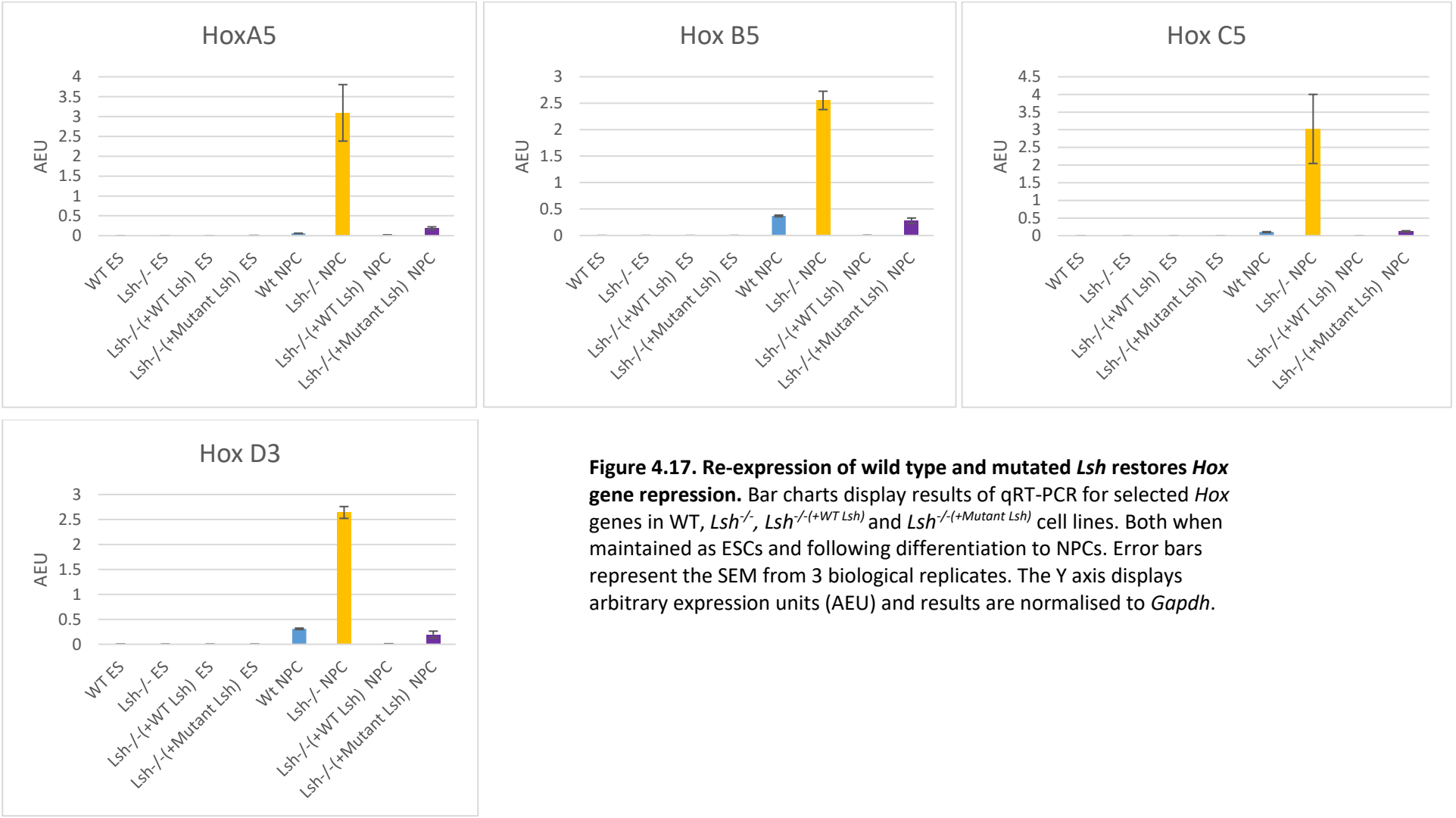


Figure 4.17. Re-expression of wild type and mutated *Lsh* restores *Hox* gene repression. Bar charts display results of qRT-PCR for selected *Hox* genes in WT, *Lsh*$^{-/-}$, *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines. Both when maintained as ESCs and following differentiation to NPCs. Error bars represent the SEM from 3 biological replicates. The Y axis displays arbitrary expression units (AEU) and results are normalised to *Gapdh*.

4.3 Discussion

The aims of the work presented in this chapter were to examine the *Lsh*$^{-/-}$ DNA methylome in NPCs in more detail by ERRBS and also to examine any transcriptional consequences of the global hypomethylation that occurs upon neurodifferentiation via RNA-Seq. My discovery that many of the differentially expressed genes were involved in developmental processes led onto the examination of the Polycomb system.

4.3.1 *Lsh* is required for DNA methylation across repeats and single copy gene promoters upon neurodifferentiation, but does not influence transcription at these sites

ERRBS confirmed a requirement for *Lsh* in appropriate DNA methylation at all repetitive elements included in the analysis, namely full length Satellites, LINE1s and LTRs.

Interestingly, this analysis revealed hypomethylation at repeat elements even in *Lsh*$^{-/-}$ ES cells, which was not detected by global methylation analysis in chapter 3. This may point towards *Lsh* playing a minor role in maintenance methylation at this stage, this could be tested by repeating the analysis after multiple passages. It could, however, be argued that the slight degree of hypomethylation seen, may be due to a population of primed cells as ES cells in culture cycle between a primed and naïve state. Irrespective of this, the key role of *Lsh* in DNA methylation was during differentiation to NPCs, when *Lsh*$^{-/-}$ cells achieved only 57.1%, 64.4% and 54.1% of WT methylation at Satellites, LINE1s and LTRs respectively.

Supporting the “established” role for *Lsh* in *de novo* methylation, LINE1s gained methylation upon neurodifferentiation but to a lesser extent than in WT cells. However, ERRBS demonstrated a reduction in DNA methylation in *Lsh*$^{-/-}$ cells when differentiating from ESCs to NPCs at LTRs and Satellites. This may suggest *Lsh* is playing a role in maintenance methylation at these sites during development. Although one cannot exclude the possibility of these sites being demethylated at some point during the differentiation process. The role of *Lsh* in regulating DNA methylation therefore appears to be complex, being dependent on both genomic and cellular context. Despite this severe hypomethylation, there was no up-regulation of repeat transcription in *Lsh*$^{-/-}$ NPCs.

As there have been several reports of *Lsh* playing a role in methylation at some single copy loci, I also examined methylation at gene promoters and found a significant number of these (395) with reduced methylation. However, when examining these sites further, I noted that these hypomethylated promoters were more likely to be in regions of the genome enriched for LINE1, and to a lesser extent LTR elements, when compared to promoters that were not

differentially methylated. Although not conclusive, this would lend support for the main targets of *Lsh* being repeat elements and that changes in methylation at single copy loci may simply be an effect of their proximity to retrotransposons.

4.3.2 There is misregulation of developmental gene expression in *Lsh*$^{-/-}$ NPCs

Although a number of genes were differentially expressed in *Lsh*$^{-/-}$ cells, particularly upon differentiation, the trend was not towards an overall up-regulation which might be expected in a hypomethylated model system. Indeed there was no correlation between promoter methylation and transcriptional state for the vast majority of genes. Those genes that did show up-regulation were not in repeat rich regions of the genome.

Differentially expressed genes mainly had functions involved in developmental processes. This prompted me to investigate Polycomb regulation in my system by carrying out ChIP for the repressive H3K27me3 modification which is deposited by PRC2.

4.3.3 *Lsh* is required for appropriate deposition of H3K27me3

ChIP experiments revealed a redistribution of H3K27me3 away from promoter sites to hypomethylated repeats. Gene ontology terms of promoters losing this Polycomb mark in *Lsh*$^{-/-}$ NPCs were to do with neurodevelopmental processes. This may therefore account for the greater propensity for *Lsh*$^{-/-}$ cells to differentiate down neural lineage as demonstrated in chapter 3. Unlike promoter DNA methylation, the loss of H3K27me3 was correlated with transcriptional changes seen in *Lsh*$^{-/-}$ NPCs.

This redistribution mirrors what is seen in other hypomethylated models. The working theory being that hypomethylation provides more binding sites for Polycomb, usually mutually exclusive from DNA methylation, resulting in a dilution effect across the genome.

Re-expression of wild type *Lsh*, which rescued the methylation defect, also rescued H*ox* gene repression. However, somewhat unexpectedly, *Hox* gene repression, but not DNA methylation, was rescued in *Lsh*$^{-/-}$ cells re-expressing a mutated version of *Lsh*. This suggests something more complex than a simple redistribution of H3K27me3 to hypomethylated sites causing an up-regulation of Polycomb target genes is occurring.

Possibilities include: firstly, that *Lsh* may play a direct role in targeting Polycomb, at least to *Hox* genes. Secondly it may be that *Lsh* is playing a role in regulating the transcription of developmental genes via interaction with other epigenetic modifiers for example via the intact N-terminal coiled coil domain, independent of any effects it has on DNA methylation. This does not argue against a redistribution hypothesis due to hypomethylation across the genome as a whole, but argues that *Lsh* may have additional roles at developmentally important genes.

In the future it would be interesting to further explore these hypotheses with ChIP-seq experiments for H3K27me3 and the LSH protein in the *Lsh*$^{-/-(+WT\ Lsh)}$ and *Lsh*$^{-/-(+Mutant\ Lsh)}$ cell lines.

So far the work I have presented has focused on the requirements for *Lsh* very early in neurogenesis in a cell line model which, whilst very informative, does not necessarily reflect the role of *Lsh in vivo*. In Chapter 5, I will discuss the use of a mouse model to investigate the role of *Lsh* at a later stage in neurodevelopment and the consequences of the absence of *Lsh* during development upon the mature brain.

Chapter 5. The use of a novel targeted knockout mouse model to investigate the role of *Lsh in vivo*

5.1 Introduction

In chapters 3 and 4, I demonstrated the requirements for *Lsh* in early neurogenesis i.e. from ES to neural progenitor cell stage using a cell culture model. Whilst a very useful model system, it does not necessarily reflect the role *Lsh* plays in DNA methylation during neurodevelopment *in vivo.* Additionally, one cannot use this system to investigate the consequences of *Lsh* loss during development on the mature adult brain. Furthermore, it is unknown if *Lsh* also has a role later in neurodevelopment as NPCs differentiate to mature neurons.

As previously described, *Lsh* knockout mice die within a few hours of birth, or in the case of a hypomorphic model 40% can survive up to several weeks postnatally with many comorbidities (Geiman et al., 2001; Sun et al., 2004). This limits their role as models for studying the longer term effects of any developmental DNA methylation abnormalities. Furthermore, the effects solely on neurodevelopment and functioning are difficult to assess when there are failures in multiple body systems. Similar difficulties are found in *Dnmt* knockout mouse models which are lethal in embryonic or early postnatal stages. Even neurally targeted *Dnmt* knockout models are either lethal, show no hypomethylation or are only targeted to a subpopulation of neurons (Fan et al., 2001) (Golshani et al., 2005) (Hutnick et al., 2009) (Nguyen et al., 2007) (Feng et al., 2010a).

Our hope was that by using a neurally targeted *Lsh* knockout mouse model, we would have a novel model, that exhibited methylation defects during development, and would survive to adulthood without comorbidities in body systems outside of the central nervous system. Such a mouse was generated by Dr Ian Adam's laboratory. Using this model, I aimed to assess the requirement for *Lsh* in DNA methylation at a later stage of neurodevelopment. As a *Nestin-Cre* model, *Lsh* will not be deleted until NPC stage. I also wished to examine any consequences upon the mature brain. This has not been feasible with previous models and is of great import given the recent discovery that mutations in *Lsh* can be causative for ICF syndrome in which patients suffer from intellectual disability of unknown pathogenesis.

5.2 Results

5.2.1 Characterisation of a targeted *Lsh* knockout mouse model

5.2.1.1 Genotype of pups resulting from final cross

The generation of the mouse model is described in section 2.3.1. The scheme essentially created a targeted deletion in neural progenitor cells of exon 12 of *Lsh*, which encodes the conserved helicase domain IV. Table 5.1 displays the genotype of pups across a sample of litters from the final cross (*Lsh*$^{fl/+}$ *Nestin Cre x Lsh*$^{fl/+}$) in the generation of the mouse model.

Genotype	***Lsh*$^{+/+}$**	***Lsh*$^{+/fl}$**	***Lsh*$^{fl/fl}$**
Cre^{+}	4 (10.5% vs 12.5%)	4(10.5% vs 25%)	7(18.4% vs 12.5%)
Cre^{-}	5 (13.2% vs 12.5%)	11(29.0% vs 25%)	7(18.4% vs 12.5%)

Table 5.1. Genotype of pups. The table shows the genotype of pups across a sample of litters from the final cross (*Lsh*$^{fl/+}$ *Nestin Cre x Lsh*$^{fl/+}$) in the generation of the mouse model. Displayed are the total number of pups with each genotype and in brackets the percentage of pups with this genotype versus expected percentage according to Mendelian inheritance. *Lsh*$^{+}$ refers to the wild type allele and *Lsh*^{fl} to the transgenic floxed allele.

Average litter size was 7 pups which is in keeping with the average mouse litter size of 6-8 signifying no high rates of embryonic death. Males and females were born at the expected ratio (42.9% and 57.1% respectively). Although genotyping of the pups was not quite at Mendelian frequency this is likely due to the fact that, by necessity, this analysis was done over a relatively small sample size of 38 pups. The fact that there were normal rates of pups homozygous for the transgenic allele and carrying *Cre* rules out any embryonic lethality due to the transgene itself. From this point on, mice with the genotype *Lsh*$^{fl/fl}$ *Cre*$^{+}$ will be referred to as the knockout model (KO). Control mice are of the genotype *Lsh*$^{fl/fl}$ *Cre*$^{-}$. Only males were used in all analyses.

5.2.1.2 Verification of Recombination

Nestin expression occurs in the developing mouse brain at around E10 reaching a peak globally at E14-E15 although different brain regions will show different expression levels according to areas of ongoing neurogenesis (Mignone et al., 2004). Nestin expression is down-regulated as neural differentiation reaches its terminal stages. In order to verify *Cre-Lox* recombination I carried out PCR genotyping on different tissues derived from KO and control 12 week old mice using the primer pair illustrated in Figure 5.1 spanning the floxed exon 12. Representative results are shown in the agarose gel in Figure 5.1. The smaller

band (820 bp) on the gel represents a recombined allele, i.e. a knockout sequence whereas the larger band (1540 bp) represents a non-excised floxed, i.e. wild type, sequence. This verified successful recombination of *Lox-P* sites in the *Lsh* gene within target tissues of forebrain and cerebellum as consistently in all mice tested there was no larger band visible on the gel. There was also some evidence of partial recombination in other tissues such as kidney and spleen as has previously been well described in *Nestin-Cre* mice (Dubois et al., 2006).

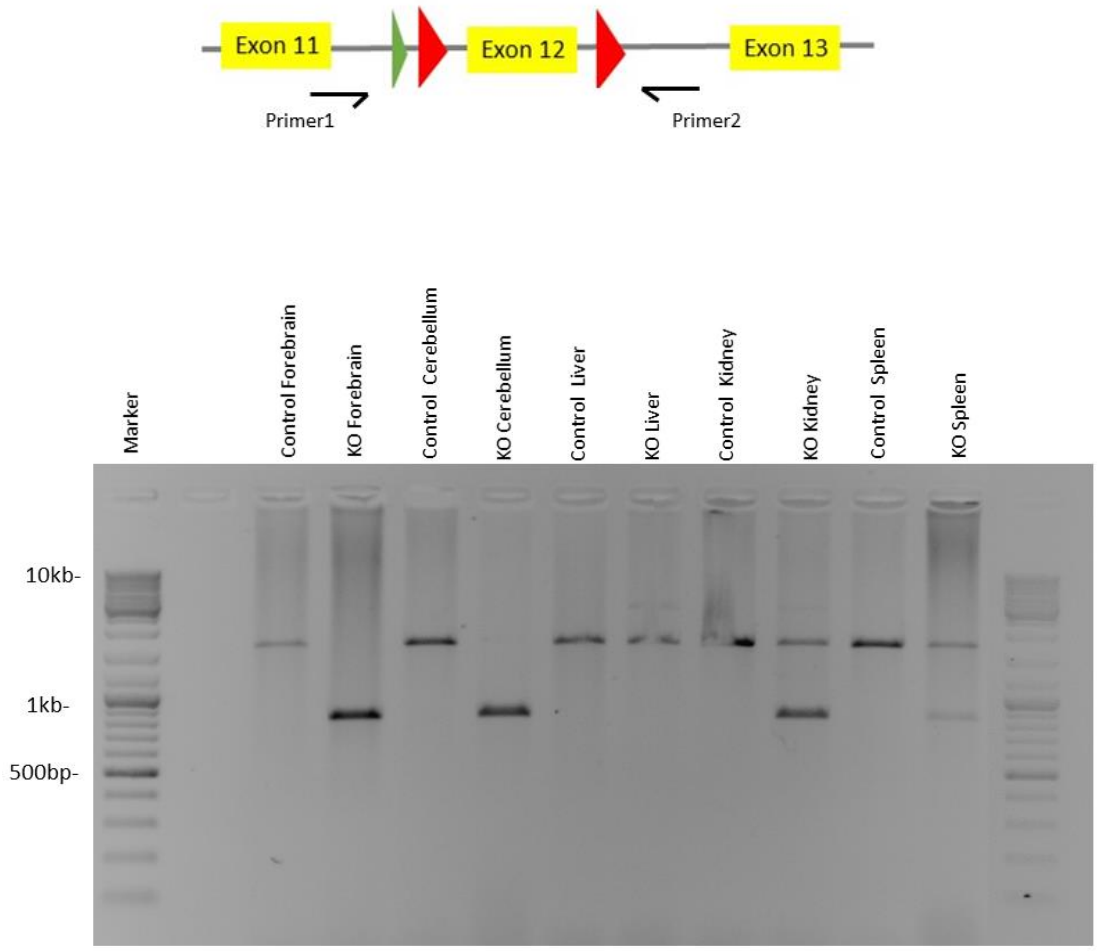
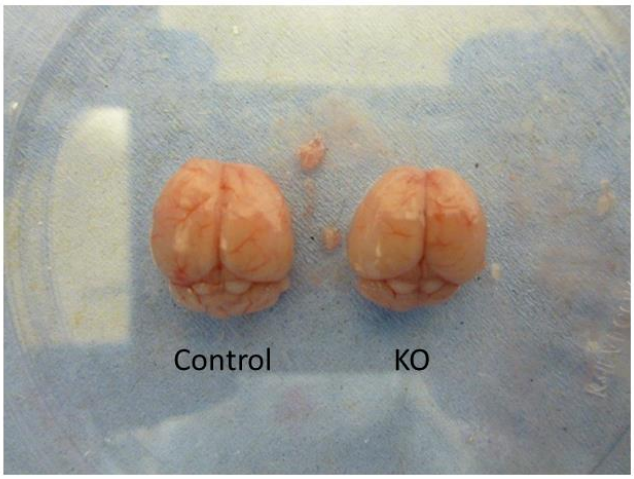


Figure 5.1. Verification of recombination. A schematic of the primer pair used to verify recombination in KO target tissues spanning the floxed exon 12 of the *Lsh* gene and a representative agarose gel from various tissues in KO and control mice. Successful recombination (shown by the presence of a smaller band) is shown in knockout mouse forebrain and cerebellum with a lesser degree of recombination also occurring in non-target tissues including kidney and spleen.

5.2.1.3 Gross brain morphology

As regulation of developmental gene expression and neurodifferentiation was perturbed in *Lsh*$^{-/-}$ cells one might hypothesise that if this was occurring *in vivo,* there may be alterations in brain morphology due to the precise nature of neurogenesis timing. For example, premature neural differentiation could result in reduced brain weight, reduced cortical thickness and a reduction of the adult neural stem cell pool.

The brains of 12 week old mice (at which point the brain is considered “adult”) were examined. This revealed no significant difference in total brain weight between control and KO mice (Figure 5.2).



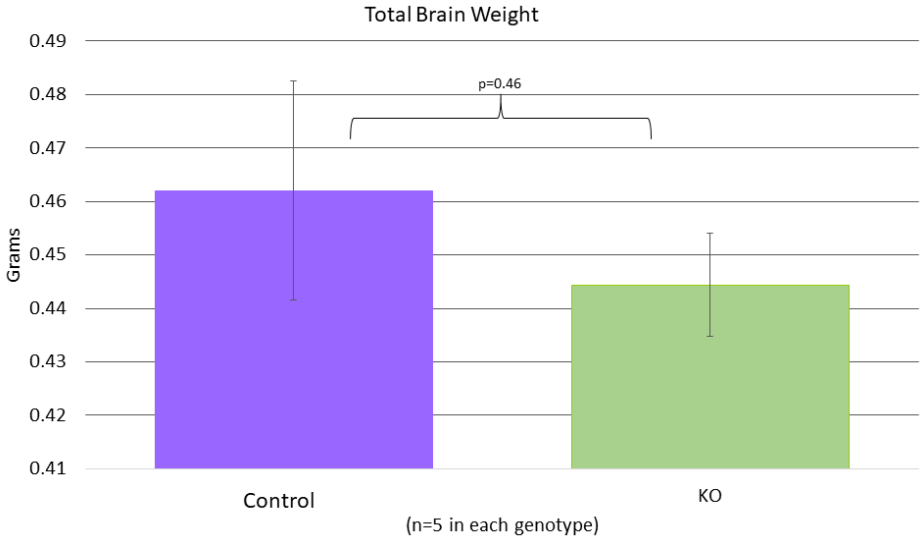


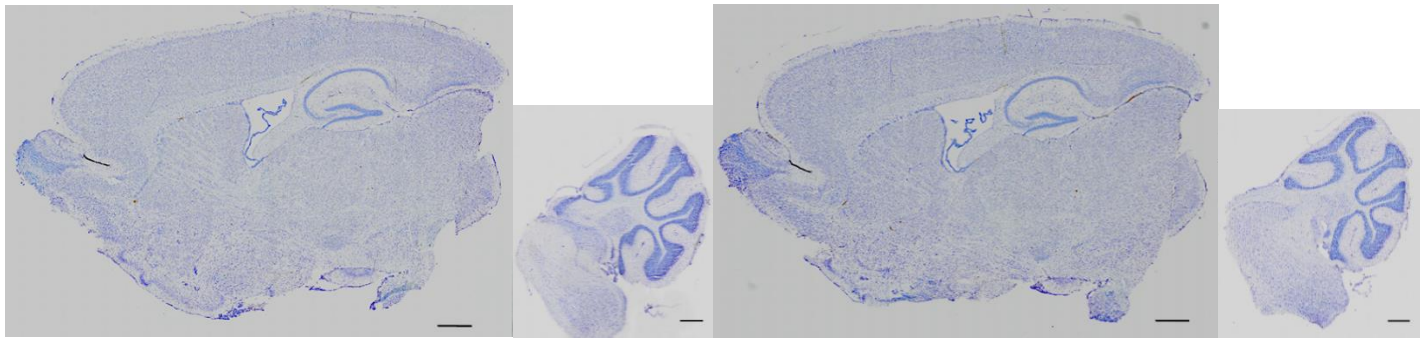
Figure 5.2. Lsh *knockout* mouse brains are of normal size. Image of control and KO mouse brains. Bar chart displaying no difference in total brain weight. Error bars represent SEM from measurements from 5 mice. P-values indicate level of significance calculated by two-tailed student t-test.

Neuroanatomy was analysed using cresyl violet staining, which stains Nissl substance in the cytoplasm of neurons, revealing no anatomical abnormalities in the knockout mouse model (Figure 5.3 A). Brain sections were also reviewed by consultant neuropathologist Dr Colin Smith, who confirmed no gross changes in morphology. Measurements were also taken of specific brain areas including somatosensory and somatomotor cortical thickness, hippocampal area and the area of the granular layer of the dentate gyrus (the site of the adult neural stem cell pool) which again revealed no significant differences between control and knockout models (Figure 5.3 B). Although no gross morphological differences were found, it would be interesting, in the future, to investigate this using more sophisticated means such as staining for neural stem cells and for markers specific for each cortical layer with larger numbers of mice.

In the *in vivo* situation there are obviously other external factors, such as diffusible signalling molecules, driving neurogenesis which may explain why we do not see alterations in brain morphology hypothesised from cell culture work where neurogenesis is largely driven by autocrine signalling (Hoch et al., 2009). Another reason we may not see morphological changes are that in this knockout mouse model, *Lsh* has not been deleted until the NPC stage. Therefore, any consequences *Lsh* has upon developmental gene expression and subsequent neurogenesis prior to neural progenitor cell stage may be missed in this model. In saying this, no gross changes in brain morphology have been described in models where *Lsh* is absent from the very beginning of development (Geiman et al., 2001; Sun et al., 2004).

Therefore *Lsh* appears to be dispensable for normal brain morphological development.

A Control KO



B

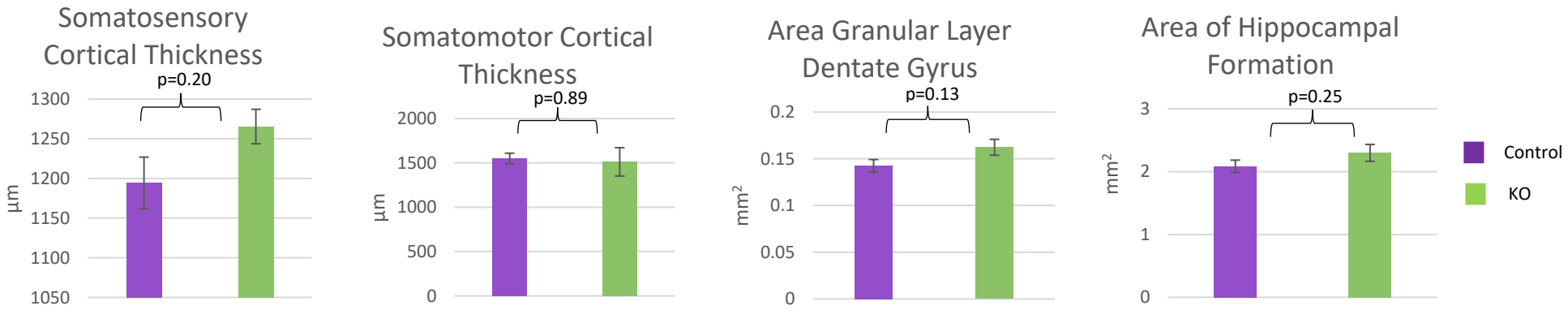


Figure 5.3. *Lsh* knockout mouse brains show no alterations in gross morphology. A. Cresyl violet staining of mouse forebrain and cerebellum revealed no gross anatomical abnormalities in the knockout model. Scale bars=500µm **B.** Bar charts display the average measurements of the somatomotor and somatosensory cortical thickness, area of the granular layer of the dentate gyrus and hippocampal area. No significant difference (tested by two-tailed student t-test) between control and knockout mice were found. Error bars represent SEM from 27 measurements in total (3mice x 3 sections x each measurement repeated 3 times).

5.2.1.4 Behavioural Analysis

A large proportion of ICF patients have intellectual disability, including those with causative mutations in *Hells* (Alghamdi et al., 2018). Therefore, I was keen to learn if our mouse model showed any deficits in behavioural testing. Furthermore, other hypomethylated mouse models, namely knockouts of *Dnmts* targeted to specific neural subtypes, show deficits in learning and memory and motor abnormalities (Feng et al., 2010a; Hutnick et al., 2009; Nguyen et al., 2007). Given limited mouse numbers with which to carry out the behavioural analyses the type of tests that could be performed was restricted. I reasoned, however, that if there were gross motor abnormalities or memory defects that these would be detectable with open field and spontaneous alternation tests. Three one year old mice from each group were analysed.

The open field test is one of the most commonly used tests of animal behaviour, allowing measurements of both general locomotor activity and of anxiety-like behaviours. Ambulatory behaviour of the mice was measured by counting line crossings. Anxiety-like behaviours were also measured by recording the time spent in the centre square and crosses into the centre square. These recordings were used as the degree of thigmotaxis (which is the tendency to remain close to the outer walls) has been validated as a measure of anxiety in mice, with thigmotaxis increasing as anxiety increases (Simon et al., 1994). I reasoned any gross motor abnormalities such as a cerebellar ataxia should be detected in an open field test even with a small number of test subjects. The anxiety measures are likely underpowered, I nevertheless felt it was important to display these results as the measures of anxiety and locomotion are clearly intertwined and can impact on each other. Recordings were taken over 2 separate tests as this also allows one to assess novel environment exploration and habituation.

It would be really interesting to assess learning and memory in the knockout mice. Behavioural paradigms used to evaluate this typically require large numbers of test subjects. With the limited numbers of mice available, I reasoned that it would only be suitable to assess working memory by spontaneous alternation in the y maze as this should be subject to less variability than other behavioural paradigms such as novel object recognition or the Barnes maze test. This test allows assessment of both working memory and exploratory behaviour. The test subject is placed in the maze and allowed to explore the three arms freely. Mice with intact cognition should show a preference to explore a less recently visited arm. This can be quantified by calculating the percentage of spontaneous alternation which is the number of successful triads (i.e. each of the 3 arms entered in turn) divided by the total number of possible triads.

No significant differences were found in all measures in the open field and y maze tests (Figure 5.4 A&B). One can conclude from this that there were no differences in exploratory behaviour, general motor activity or working memory. With larger numbers of mice it would be interesting to carry out more detailed assessment of cerebellar motor coordination for example on the rota rod test and more complicated learning paradigms would be necessary to detect any deficits which may correlate to the intellectual disability seen in ICF.

Figure 5.4 A

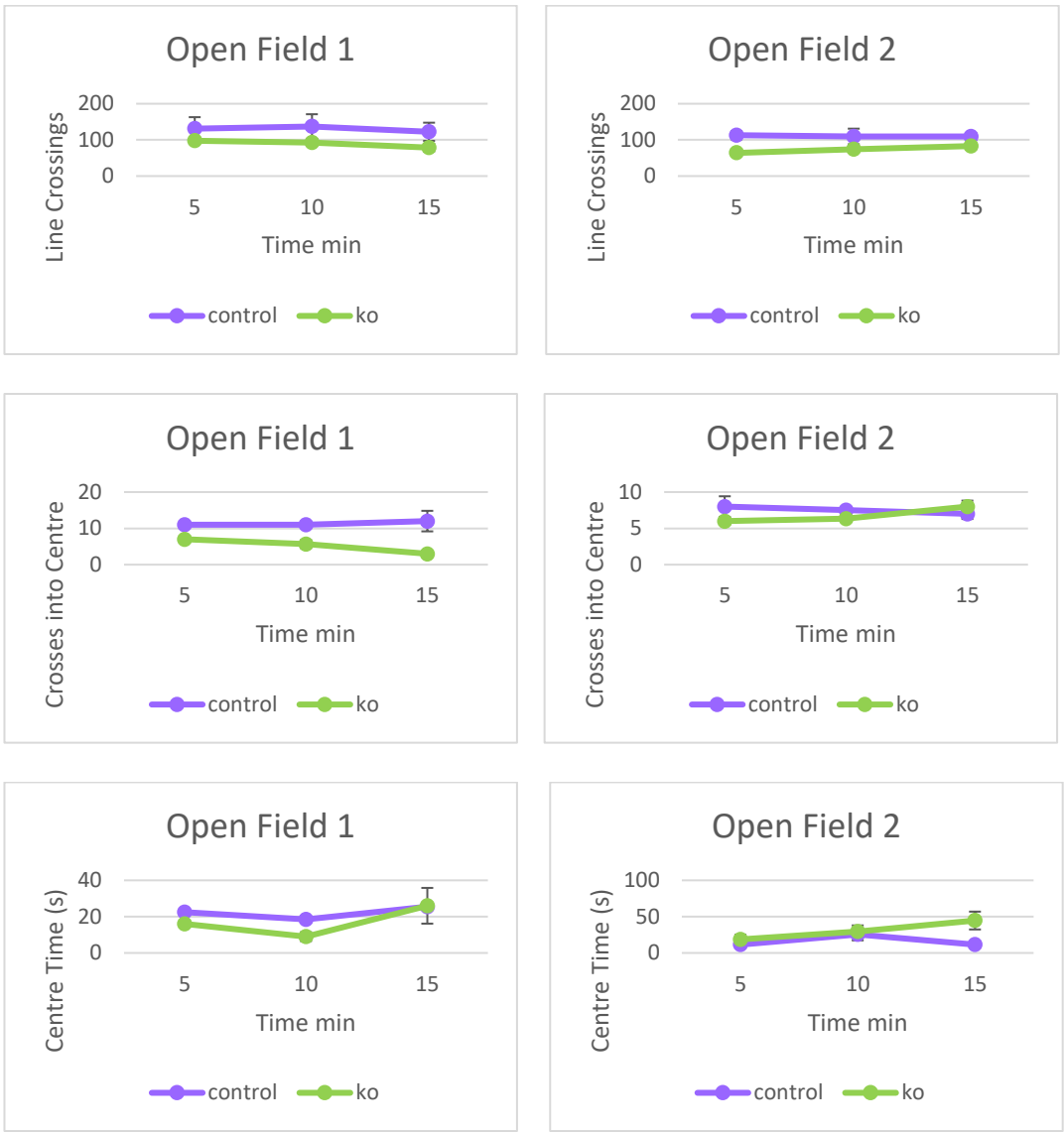


Figure 5.4. Behavioural testing in open field and Y maze. A. Recordings from 2 separate open field tests of line crossings, time spent in centre square and crossings into centre square providing information on ambulatory and anxiety-like behaviours in the mice. Results are displayed across 5 min intervals.

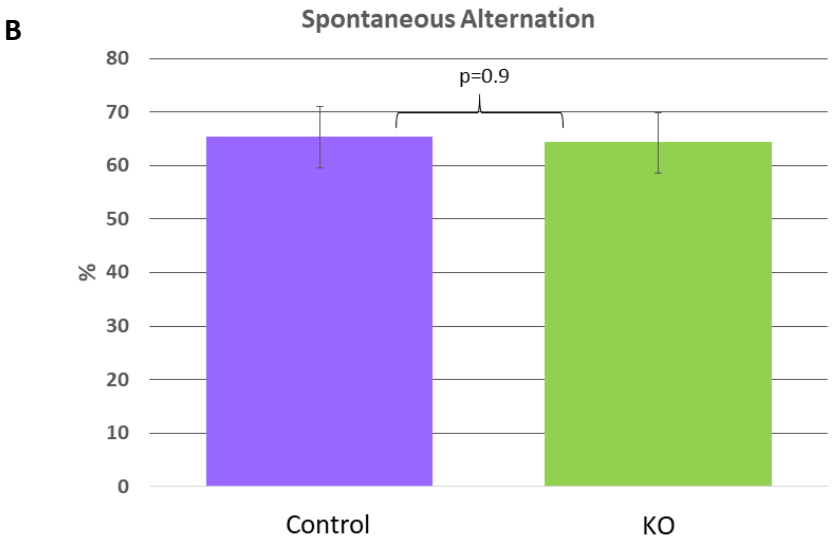


Figure 5.4. Behavioural testing in open field and Y maze. B. Bar charts display percentage spontaneous alternation in the y maze, a test of working memory. There were no significant difference in results as calculated by 2 tailed t test. N=3 in each group. Lines represent +/- s.e.m.

5.2.2 Global DNA methylation levels are reduced in *Lsh* knockout mouse brain

To compare global DNA methylation levels between the knockout and control mouse brains I employed the same methods as used for the cell culture experiments described in chapter 3. Firstly I carried out restriction enzyme digest with HpyCH4IV which cuts DNA at unmethylated ACGT sites. On the gel I observed the appearance of low molecular weight bands following digestion in the knockout mouse forebrain and cerebellum which was not seen in DNA from control mice. This demonstrated hypomethylation of satellite regions in knockout mouse neural tissue (Figure 5.5). Digests were also carried out with MspI, a methylation-insensitive restriction endonuclease which cuts at CCGG sites regardless of methylation status as a control for the digestibility of the DNA.

Given that I found some Cre-Lox recombination also occurring in non-target tissues, namely the kidney and spleen, I also screened for hypomethylation in these tissues. Using restriction digest there was no evidence of hypomethylation at satellites in these tissues.

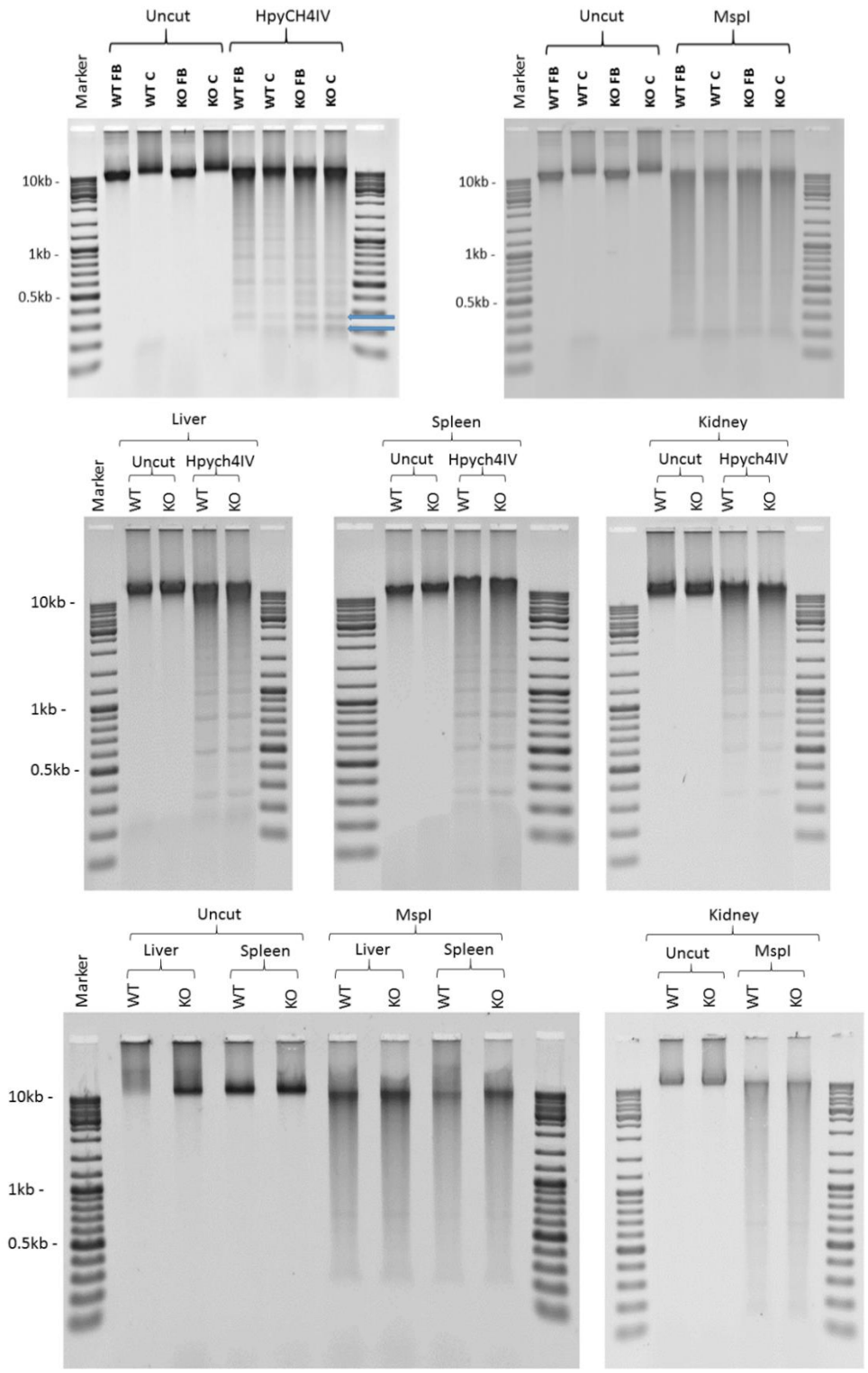


Figure 5.5. DNA methylation in *Lsh* knockout tissue A. Agarose gels displaying digestion of DNA from control and knockout mouse forebrain (FB), cerebellum (C), liver, kidney and spleen by methylation sensitive HpyCH4IV enzyme. DNA integrity and digestibility are controlled for using uncut samples and MSPI digestion respectively. Arrows indicate the presence of low molecular weight bands indicating hypomethylation of satellite sequences.

In order to better quantify global DNA methylation I then digested DNA to single nucleotides and LC-MS was performed on the samples by Jimi Wills to measure 5-mC. This demonstrated significant hypomethylation in knockout neural tissue with a ~12% reduction in methylation in knockout mouse forebrain and a~ 23% reduction in knockout mouse cerebellum (Figure 5.6). These results establish, for the first time, that *Lsh* is playing an important role in DNA methylation during neural progenitor cell differentiation to mature neurons. As *Lsh* is not expressed in mature neural tissue, this result also demonstrates that a protein expressed during early neurogenesis can have a lasting impact on the mature neuronal methylome.

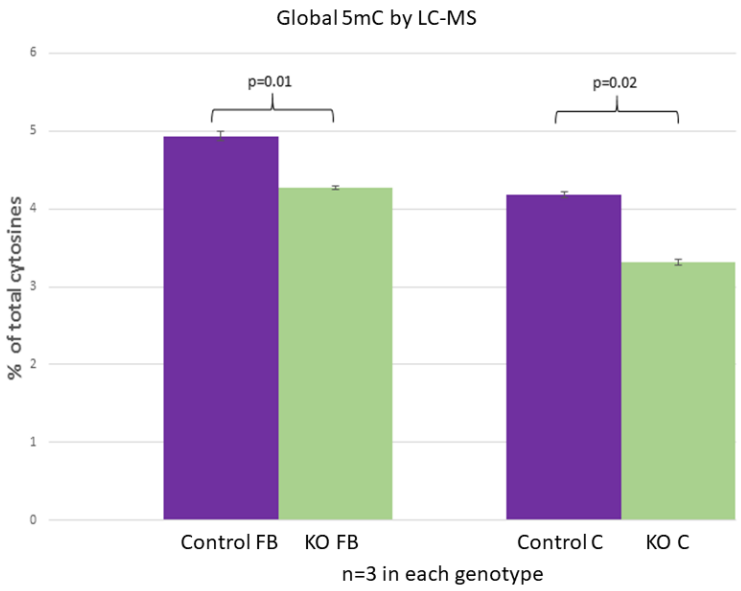


Figure 5.6. DNA hypomethylation in *Lsh* knockout neural tissue Quantification of global 5-mC levels by LC-MS given as a percentage of total cytosines in control and KO mouse forebrain (FB) and cerebellum (C). Error bars represent the S.E.M of 3 biological replicates. P-values indicate level of significance calculated by student t-test.

5.2.3 Methylation analysis by ERRBS

To interrogate the mature neural methylome further, and identify which regions of the genome had altered DNA methylation in the absence of *Lsh,* I employed ERRBS. DNA from one replicate each of KO and control forebrain and cerebellum was sent to the Epigenomics Core Facility of Weill Cornell Medicine who performed ERRBS. Bioinformatic processing was then performed by Dr Donncha Dunican.

5.2.3.1 Repeat elements are hypomethylated in the *Lsh* knockout brain

Given the finding from cell culture experiments in chapter 4, that the main role of *Lsh* appears to be methylation at repeats, I firstly looked at these regions in the tissue ERRBS data. As for cell culture experiments, to calculate repeat methylation the average methylation of individual CpGs, with a minimal depth of 5 reads, mapping to long repeats were plotted in violin plots in Figure 5.7. LINEs, LTR and Satellite repeats were all significantly hypomethylated in the knockout mouse forebrain reaching only 84.3% ($p<2.2x10^{-16}$), 77.8% ($p<2.2x10^{-16}$) and 79.2% ($p=0.01$) of control methylation levels respectively. The hypomethylation in the knockout cerebellum was far more striking with LINE1, LTR and Satellite repeats reaching only 47.8% ($p<2.2x10^{-16}$), 31.9% ($p<2.2x10^{-16}$) and 30.1% ($p=9.6x10^{-16}$) of control methylation levels respectively.

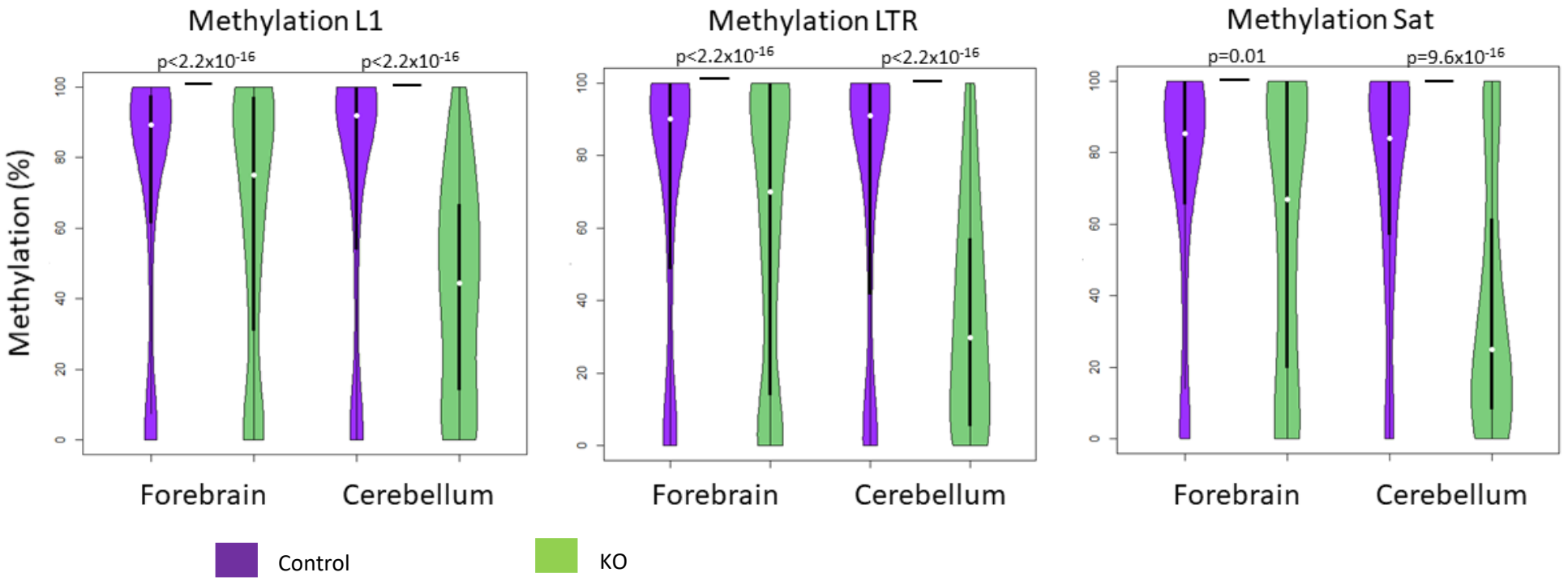


Figure 5.7. Repeats are hypomethylated in *Lsh* knockout mouse brain. Violin plots demonstrate the distribution frequency with which individual CpGs within a family of repeats are methylated in control and knockout mouse forebrain and cerebellum. White spot indicates median value and black bars represent the interquartile range. P values show results of significance testing by Wilcoxon rank sum test.

5.2.3.2 Genes associated with immune function display hypomethylation in knockout neural tissue

I was then interested to see if there would be any evidence of gene promoter hypomethylation as was found in the cell culture model. Promoter methylation was calculated as the average methylation of all CpGs spanning the +/- 2kb across the TSS and a threshold change of 20% absolute methylation difference was set to define differentially methylated promoters. This revealed hypomethylation of 164 promoters in the knockout forebrain and 99 that were hypermethylated. In knockout cerebellum a substantial 708 promoters were hypomethylated and 42 displayed increased methylation (Figure 5.8).

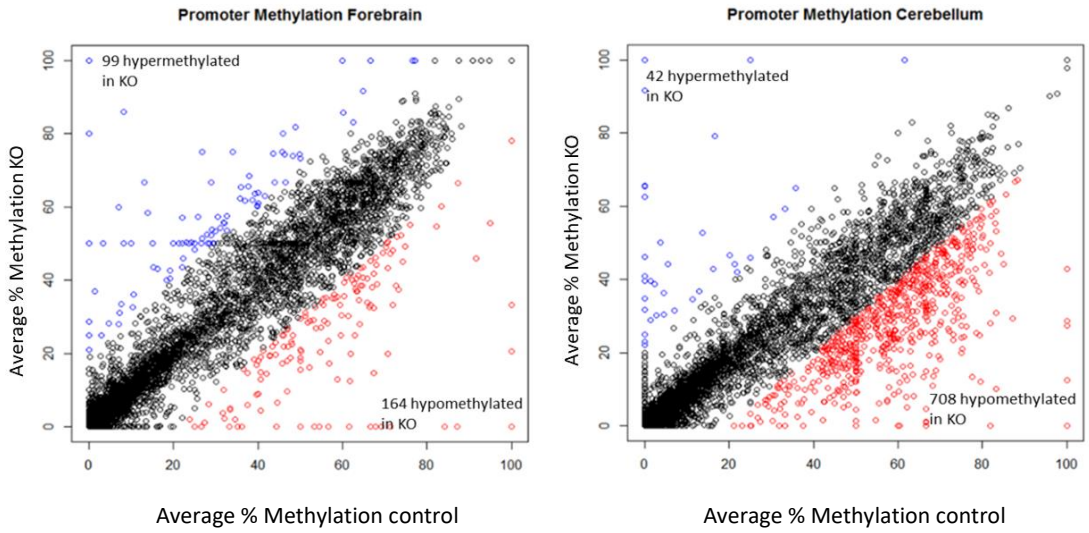


Figure 5.8. Promoter methylation in neural tissue. Scatter plots display average methylation across promoters (defined as +/- 2kb from TSS) in control and KO forebrain and cerebellum. Differentially methylated promoters (defined as >20% absolute methylation difference compared to control) are highlighted in red for hypomethylated and blue for hypermethylated in KO tissue.

Gene ontology analysis using the biological processes tool in David V6.8 revealed an enrichment of terms to do with innate immune responses (Figure 5.9 A), bearing a striking resemblance to those gene ontology terms found to be enriched in hypomethylated promoters of *Lsh*$^{-/-}$ NPCs in chapter 4. In order to assess this similarity more stringently, I overlapped hypomethylated promoter files from KO mouse forebrain, cerebellum and *Lsh*$^{-/-}$ NPCs. This revealed an overlap of 69 genes between forebrain and cerebellum and 29 genes between all 3 models (significantly more than would be expected by chance (by hypergeometric testing). This suggests that similar pathways are affected by loss of *Lsh* both in NPCs and mature neurons but that the effects of loss of *Lsh* clearly varies depending on the cellular context (Figure 5.9 B).

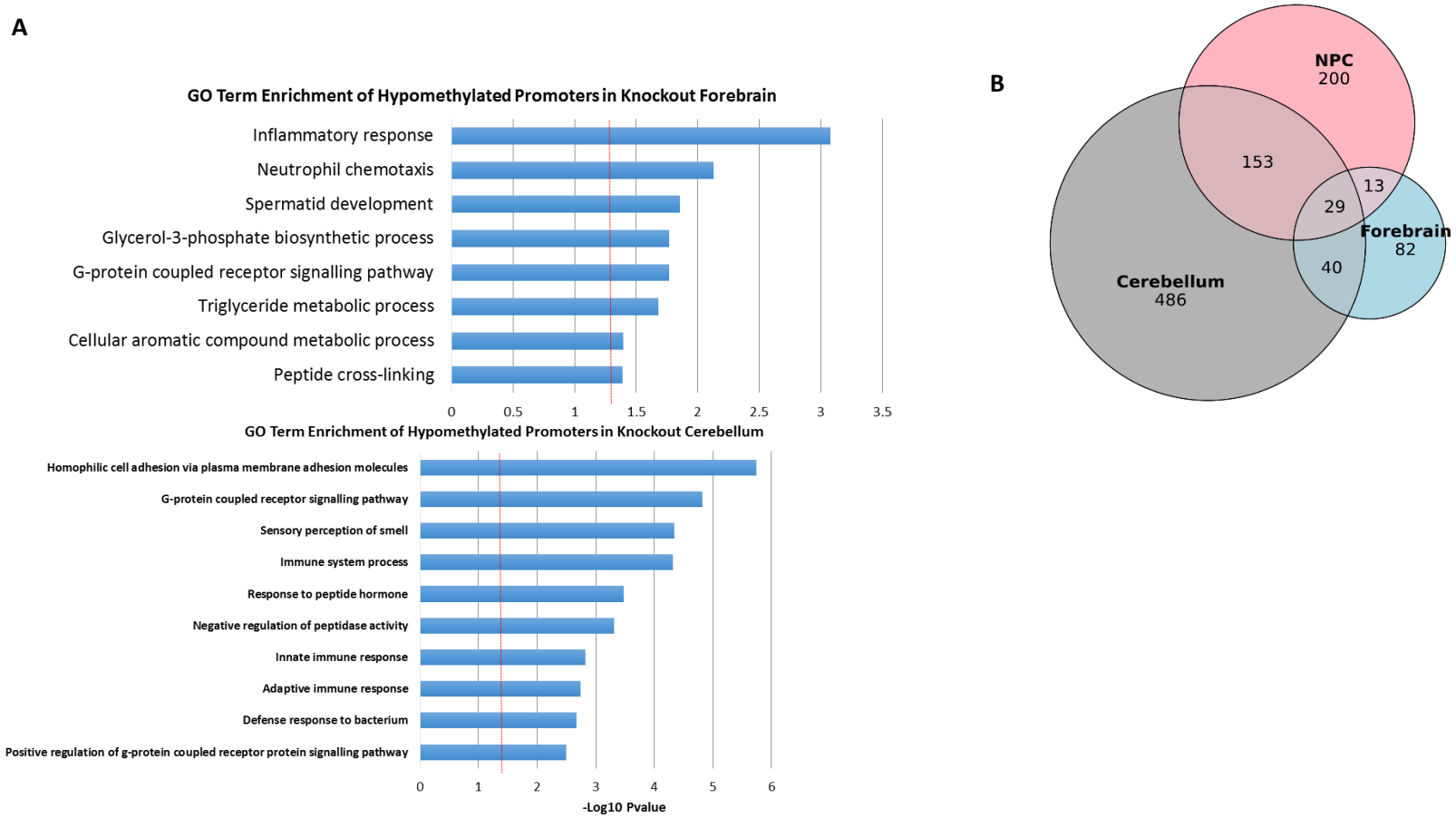


Figure 5.9. Genes associated with innate immunity are hypomethylated in KO neural tissue. A. Gene ontology analysis of hypomethylated promoters in KO neural tissue. Table shows the top most significant biological process terms. Red line demarcates p=0.05 **B.** Venn diagram displaying the overlap of hypomethylated gene promoter lists across KO mouse forebrain and cerebellum and *Lsh*$^{-/-}$ NPCs (overlapping gene lists are displayed in Appendix Lists A3-6).

5.2.3.3 L1 repeats are enriched at hypomethylated gene promoters in KO neural tissue

Given the main role of *Lsh* in methylating repeats, there exists the possibility that the promoters found to be hypomethylated are simply so because they are in the vicinity of hypomethylated repeats. This was tested using the same method as in section 4.2.1.3.

In agreement with what was seen in the cell culture model, this analysis revealed that that the 20kb stretch of genome surrounding the TSSs of hypomethylated promoters in KO neural tissue was enriched for L1 elements when compared to the regions surrounding non-differentially methylated promoters. This was also true, although to a lesser extent, for LTR content at these regions in KO cerebellum. LTR content however was not enriched at these regions in forebrain although the data had more noise due to the smaller number of hypomethylated gene promoters in this tissue (Figure 5.10).

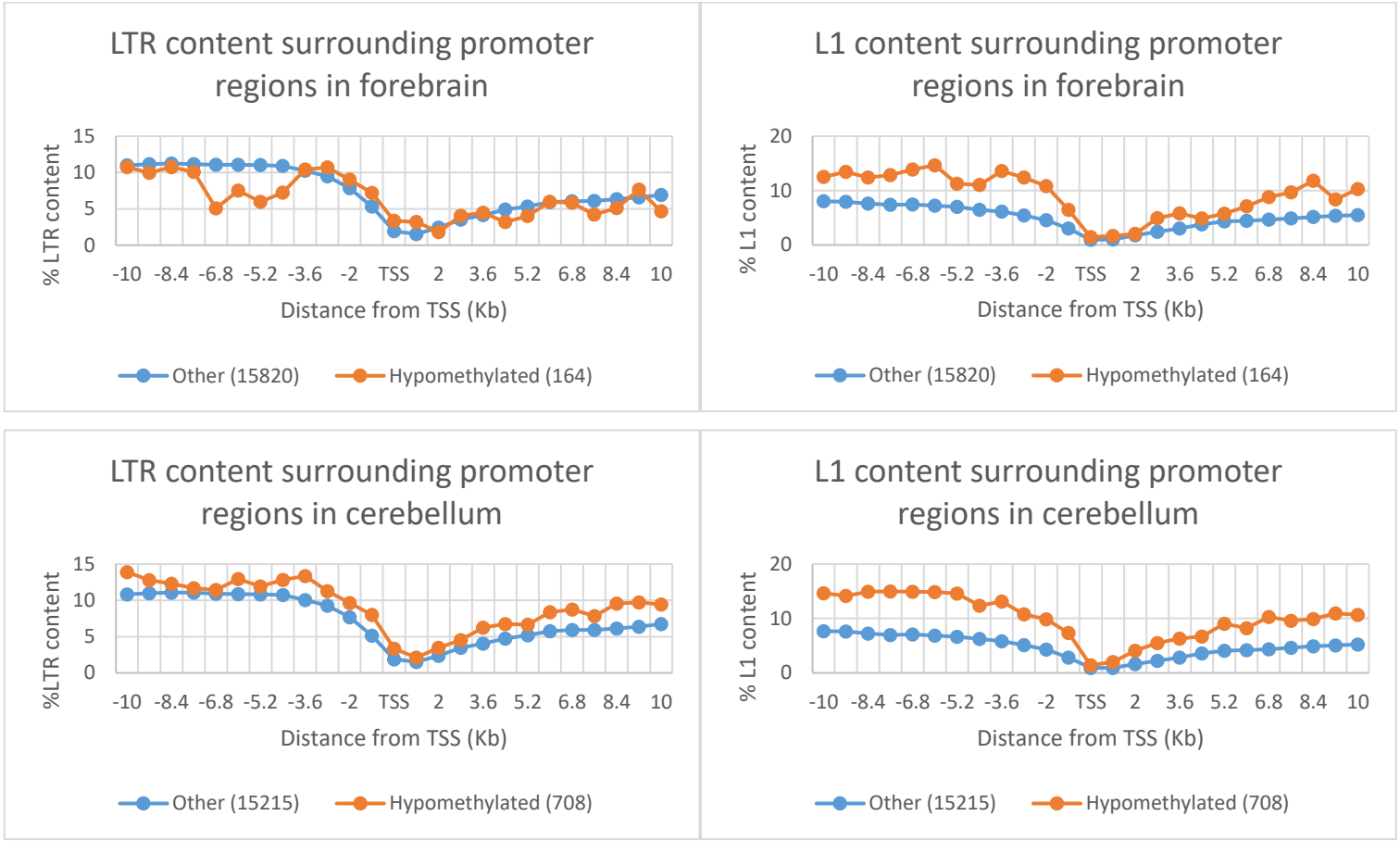


Figure 5.10. Repeat content of genome surrounding hypomethylated promoters in KO neural tissue. Graphs divide the genome surrounding the TSS of genes, with hypomethylated promoters in KO forebrain and cerebellum (orange line) and genes with unchanged promoter methylation status when compared to control (blue line), into 800bp bins 10kb up and downstream of the TSS. Each bin is represented by a point on the graph. The average percentage of base pairs within each 800bp bin that form part of an L1 or LTR annotated repeat is shown on the y axis. The number of genes in each sample is highlighted in parenthesis.

5.2.4 Analysis of the transcriptome by RNA-Seq

Given the hypomethylation seen in the knockout mouse brains, RNA-Seq was utilised to analyse if there were any transcriptional changes. Bioinformatic processing was carried out by Dr Donncha Dunican.

5.2.4.1 Repeat sequences are aberrantly expressed in the *Lsh* knockout mouse brain

Expression of repetitive elements was analysed initially. As with cell data, these were analysed separately in terms of full length (>6kb in the case of LINE1s and LTRs and >1kb in the case of satellites) and "short" repeats. In contrast to the findings in the *Lsh*$^{-/-}$ NPCs in chapter 4, this analysis did reveal aberrant repeat expression within the KO mouse model (Figure 5.11 & 12).

This deregulation of repeat suppression occurred at LTRs and LINE1s and was more marked in the cerebellum. The up-regulation of LTRs and LINE1s was confirmed by qRT-PCR which also demonstrated a huge up-regulation of major satellite repeat expression in KO cerebellum which was not evident from RNA-Seq analysis (Figure 5.13). I suspect this is due to the fact that RNA-Seq was carried out from polyA-selected libraries and that these repeats do not undergo polyadenylation. This is supported from a recent paper comparing repeat expression in human tumour tissue using both polyA-selected and total RNA-Seq which revealed that repeat elements were not properly detected in polyA selection methodologies and that this was most striking for satellites which were barely detected at all (Solovyov et al., 2018). It would therefore be beneficial to repeat analysis using total RNA-Seq.

For a read to be included in this analysis it had to map uniquely to the genome. Whilst this provides a robust dataset for analysis of an individual repeat's influence over neighbouring genomic regions, it of course means that a lot of data is lost due to conservation of sequences between retrotransposons and the fact that the same repeat may be found at multiple genomic regions. The results shown in Figure 5.11 &1 2 are therefore likely to be an underestimate of repeat differential expression and so results for non-uniquely mapped reads are shown in the appendix (Figure A3).

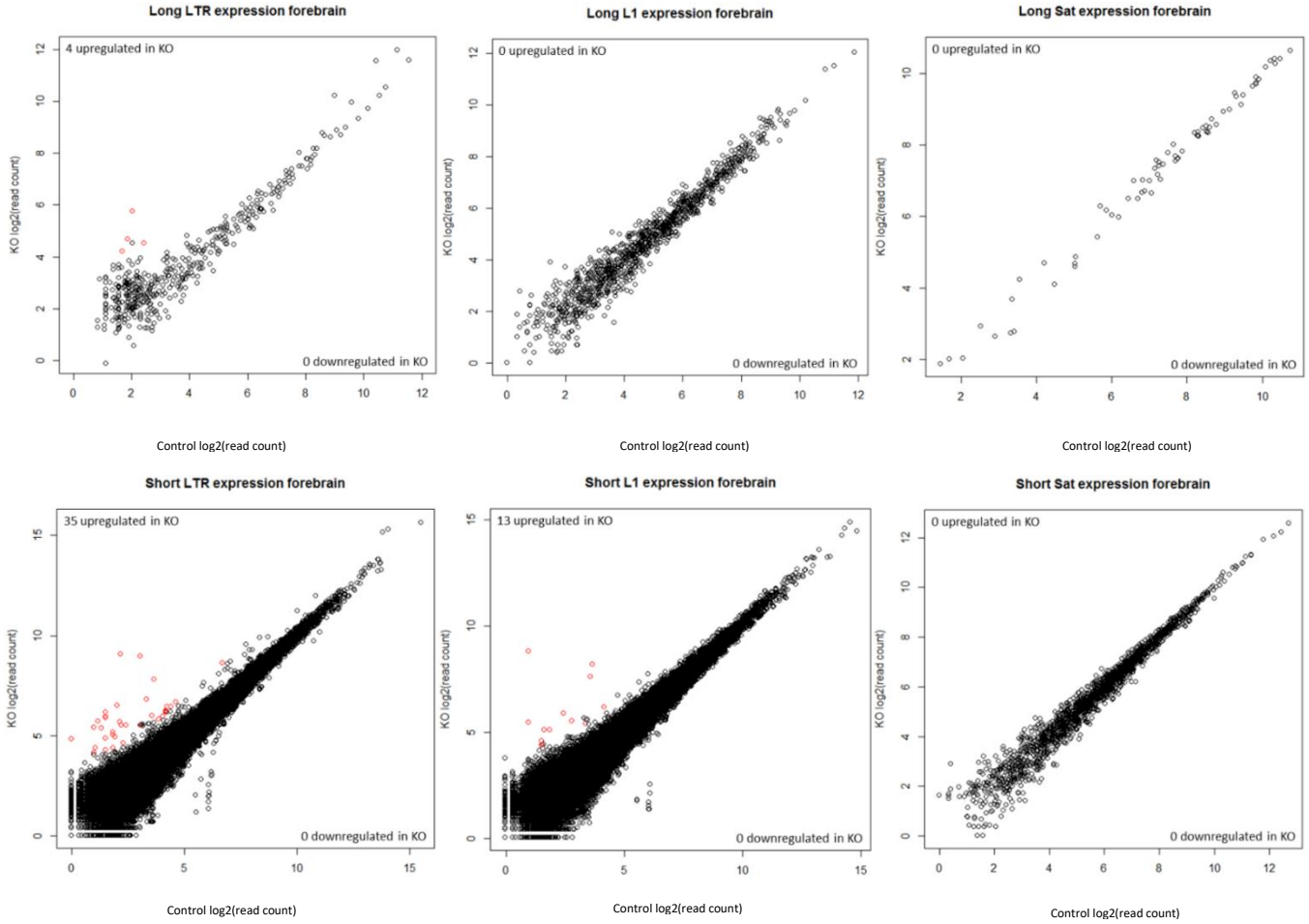


Figure 5.11. Repeat expression in forebrain. Scatter plots display read counts of LINE1 (L1), long terminal repeat (LTR) and satellite (Sat) sequences from RNA-Seq analysis from 3 biological replicates of control and KO forebrain tissue. Repeats defined as differentially expressed in KO tissue (>4 fold change and FDR<0.05) are highlighted in red if up-regulated and blue if down-regulated. Data is shown for reads mapping to full length repeats (long) and non-full length repeats (short).

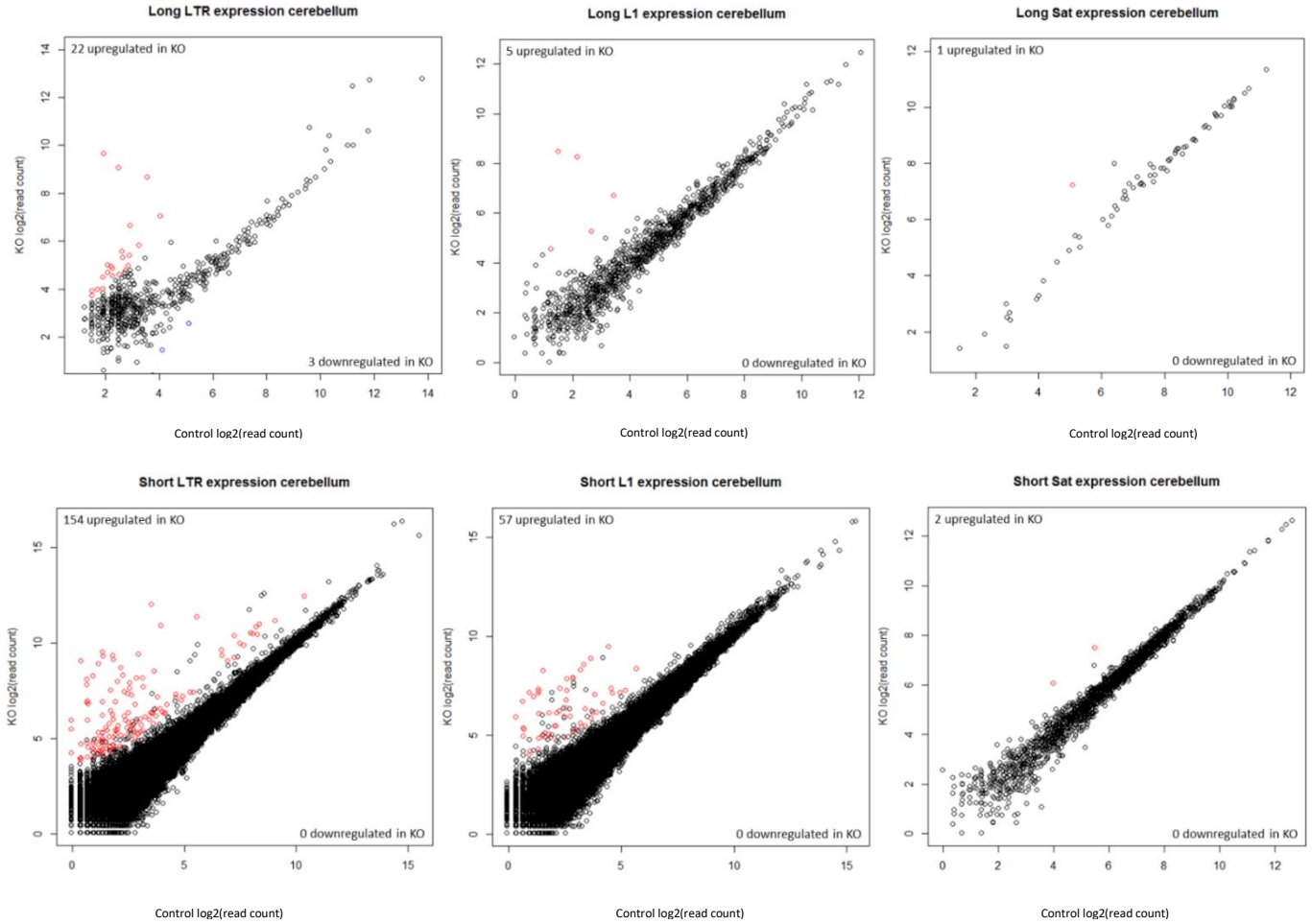


Figure 5.12. Repeat expression in cerebellum. Scatter plots display read counts of LINE1 (L1), long terminal repeat (LTR) and satellite (Sat) sequences from RNA-Seq analysis from 3 biological replicates of control and KO forebrain tissue. Repeats defined as differentially expressed in KO tissue (>4 fold change and FDR<0.05) are highlighted in red if up-regulated and blue if down-regulated. Data is shown for reads mapping to full length repeats (long) and non-full length repeats (short).

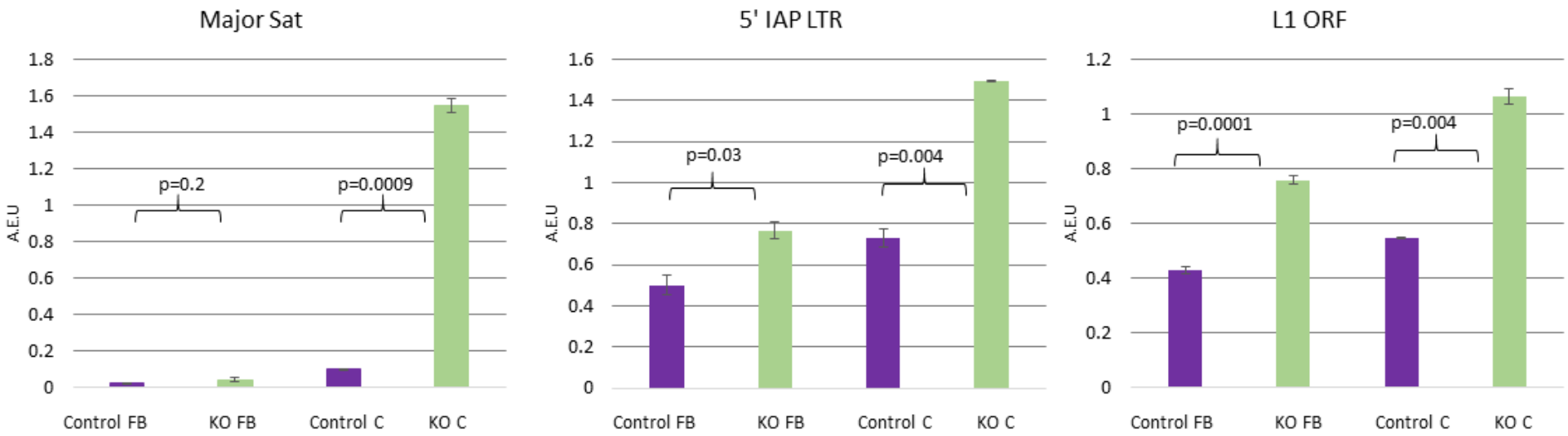


Figure 5.13. Repeat expression by qRT-PCR. Bar charts display results of qRT-PCR for major satellites, LINE1s and LTRs. The error bars represent the SEM from 3 biological replicates from control and KO forebrain (FB) and cerebellum (C). The Y axis displays arbitrary expression units (AEU) and results are normalised to *Gapdh*. Results of student t-testing are displayed on the graphs.

5.2.4.2 Genes associated with immune response are differentially expressed in the knockout mouse

Analysis of gene expression from the RNA-Seq dataset revealed fewer genes were differentially expressed in mouse knockout forebrain and cerebellum than in the cell culture experiments. Genes were considered differentially expressed if they had an >2 fold change in expression and P value <0.05 across 3 replicates. Genes with read depth <10 were excluded from the analysis. In KO forebrain 34 genes were down-regulated and 150 up-regulated and in cerebellum 41 genes were down-regulated and 224 up-regulated (Figure 5.14).

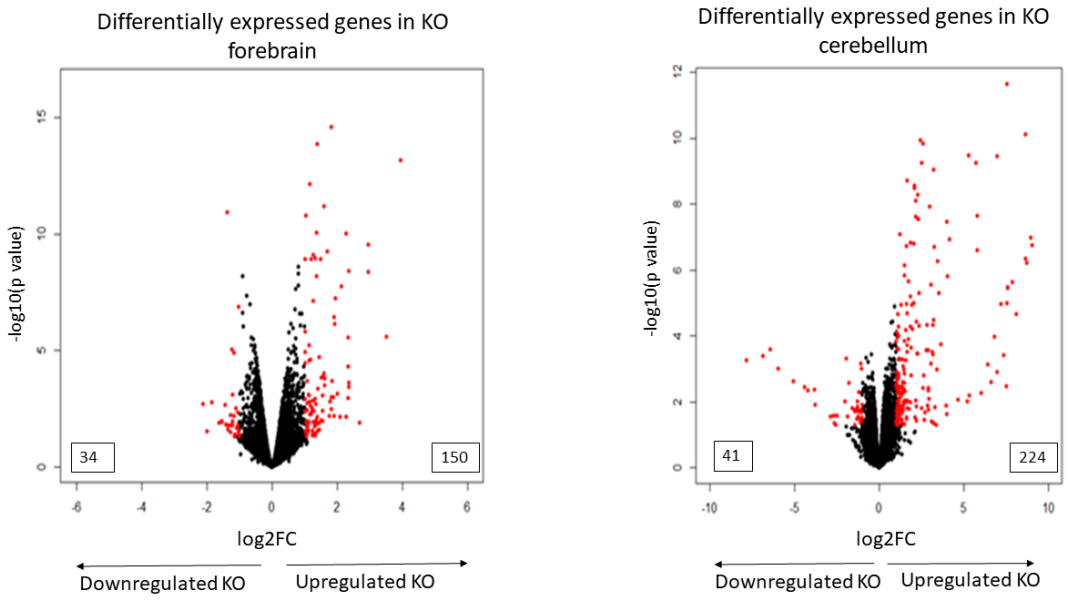
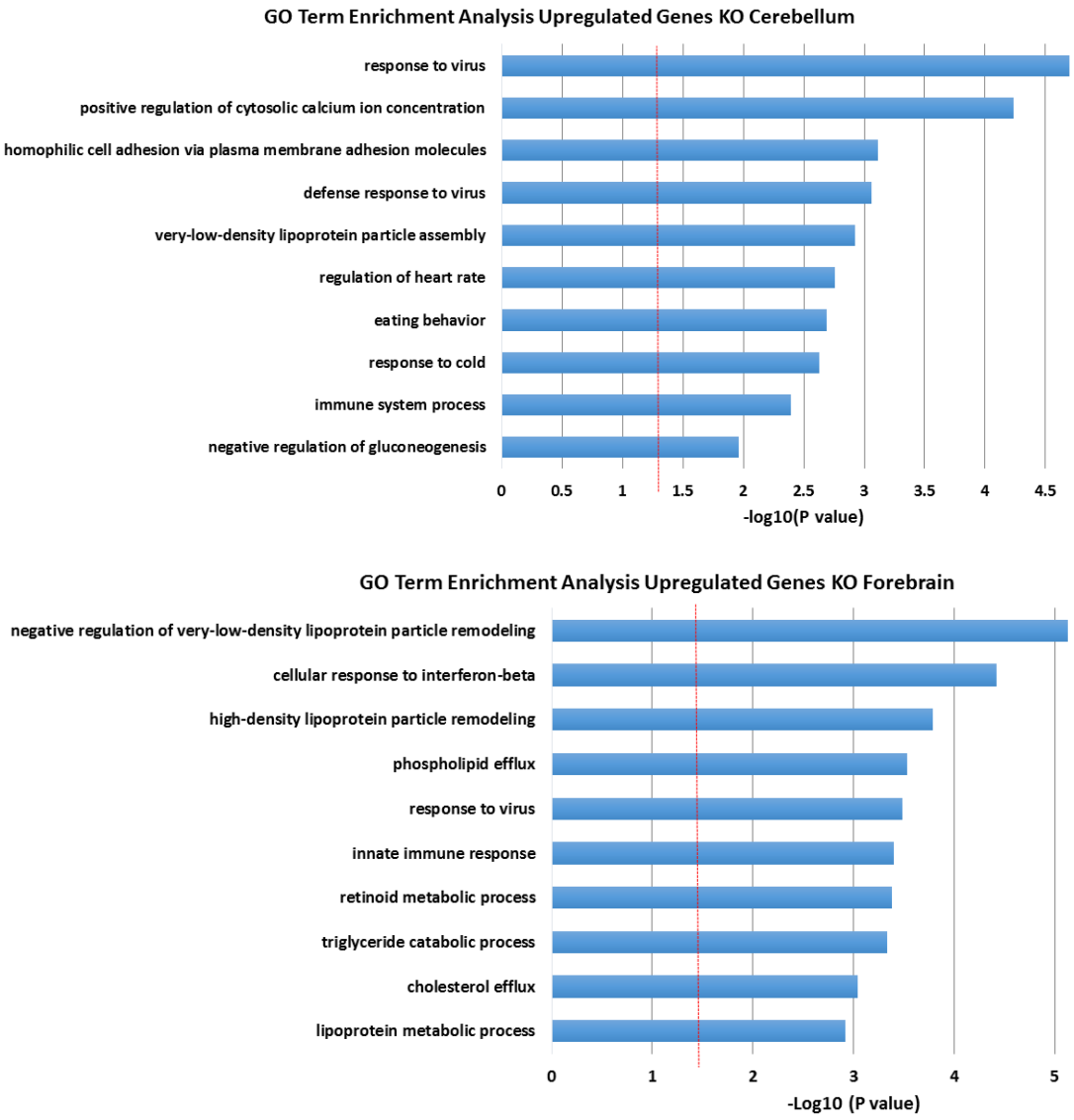


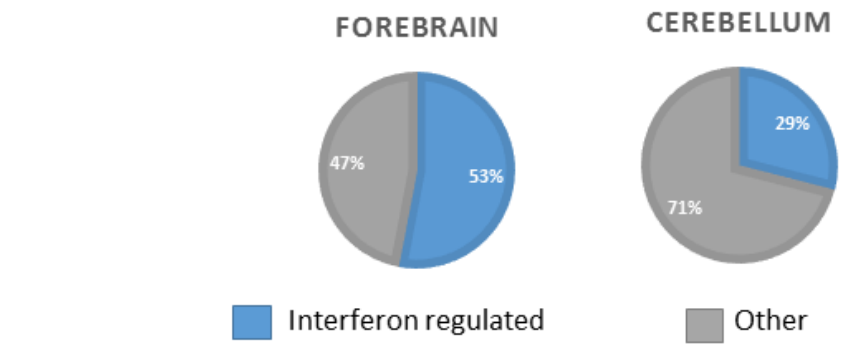
Figure 5.14. Differentially expressed genes in *Lsh* knockout mouse brain. Volcano plots display log 2 fold change in expression of genes between control and knockout forebrain and cerebellum. Genes defined as differentially expressed are shown in red with others displayed in black. Total numbers are displayed on the graph.

Up-regulated gene lists were subjected to gene ontology analysis using DAVID V6.8. This revealed a significant enrichment of terms to do with innate immune response reflecting the ontological categories enriched in the hypomethylated gene promoter lists in both *Lsh*$^{-/-}$ NPCs and KO neural tissue (Figure 5.15 A). As the number of genes in each category was limited I also uploaded the up-regulated gene lists into *Interferome V2.01,* an online tool which reports back on those genes which are known to be interferon regulated (Rusinova et al., 2013). This revealed that a large percentage of the up-regulated genes, 53% in forebrain and 29% in cerebellum, were known to be interferon regulated (Figure 5.15 B) therefore providing further evidence to validate gene ontology findings. A selection of these genes were validated by qRT-PCR (Figure 5.15 C).

Figure 5.15 A


GO Term Enrichment Analysis Upregulated Genes KO Cerebellum
response to virus
positive regulation of cytosolic calcium ion concentration
homophilic cell adhesion via plasma membrane adhesion molecules
defense response to virus
very-low-density lipoprotein particle assembly
regulation of heart rate
eating behavior
response to cold
immune system process
negative regulation of gluconeogenesis
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
-log10(P value)
GO Term Enrichment Analysis Upregulated Genes KO Forebrain
negative regulation of very-low-density lipoprotein particle remodeling
cellular response to interferon-beta
high-density lipoprotein particle remodeling
phospholipid efflux
response to virus
innate immune response
retinoid metabolic process
triglyceride catabolic process
cholesterol efflux
lipoprotein metabolic process
0
1
2
3
4
5
-Log10 (P value)

B



C

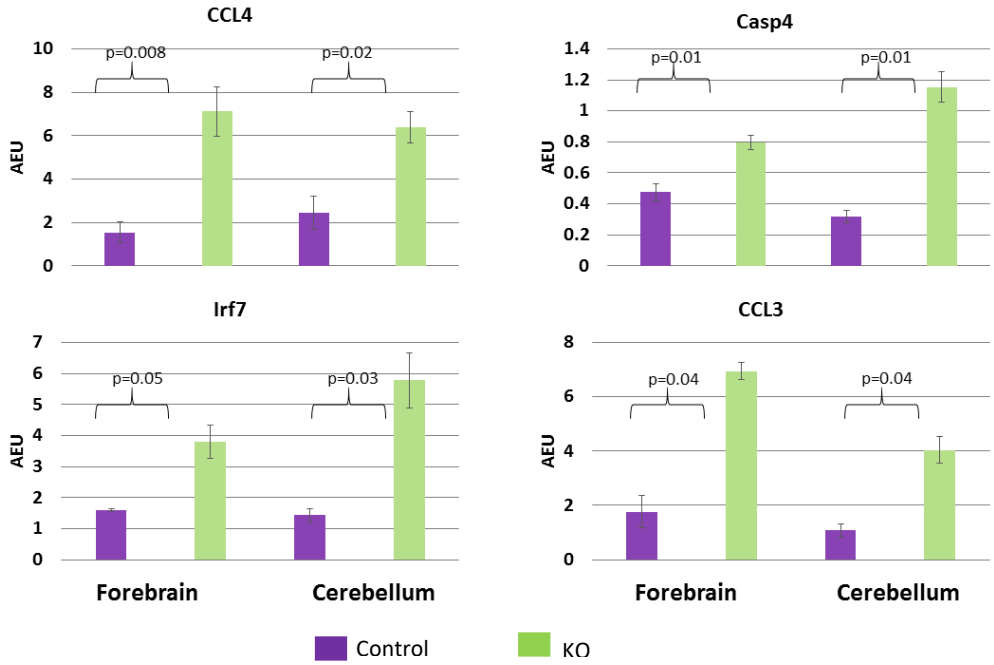


Figure 5.15. Innate immune genes are up-regulated in knockout mouse brain. A. Tables display the top ten most significant gene ontology biological process terms for genes up-regulated in KO forebrain and cerebellum. Red line demarcates p=0.05**. B.** Pie charts display the percentage of up-regulated genes that are known to be interferon regulated as analysed by the online tool *Interferome* (Lists of these up-regulated interferon regulated genes are displayed in Appendix Lists A7-8). **C**. qRT-PCR validation of expression of a selection of immune response genes found to be up-regulated in KO neural tissue. Error bars represent the SEM from 3 biological replicates. The Y axis displays arbitrary expression units (AEU) and results are normalised to *Gapdh*. P values as calculated by student t test are displayed.

In order to see if the models overlapped in their transcriptional changes I merged the up-regulated gene files of KO mouse forebrain, cerebellum and *Lsh*$^{-/-}$ NPCs. This revealed a large degree of similarity between KO mouse forebrain and cerebellum with 56 genes in common (See Appendix List A9). This was a significantly greater overlap than would be found by chance ($p=7.78x10^{-70}$), however unlike the overlap seen in the hypomethylated promoter files there was essentially no overlap with genes differentially expressed in *Lsh*$^{-/-}$ NPCs (Figure 5.16).

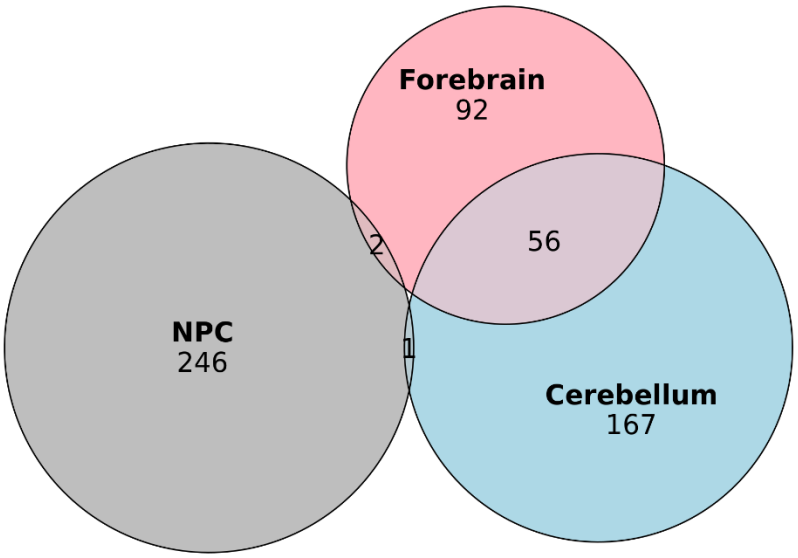


Figure 5.16 Overlap of up-regulated genes in knockout models. Venn diagram displaying the overlap of up-regulated genes across KO mouse forebrain, cerebellum and *Lsh*$^{-/-}$ NPCs.

Of particular note is the fact that up-regulated gene ontology terms from knockout tissue are devoid of developmental processes, the major terms found in the analysis of differentially expressed genes in *Lsh*$^{-/-}$ NPCs in section 4.2.2.2. The absence of key developmental genes such as *Hox* genes from the up-regulated gene list in tissue is curious given that these genes are found to be derepressed in tissues from other *Lsh*$^{-/-}$ mouse models (Xi et al., 2007). This may be due to the different timing of *Lsh* deletion between the models such that by NPC stage the majority of Polycomb target genes have been silenced for example by remodelling or DNA methylation (Liu et al., 2017; Mikkelsen et al., 2007; Mohn et al., 2008). Looking specifically at 539 genes defined as gaining H3K27me3 upon NPC to terminal neuron differentiation by Mohn *et al* (Mohn et al., 2008) only 7 were found to be upregulated in the KO cerebellum and 6 in KO forebrain. Another possibility is that this model is a hypomorph and a truncated protein is present that, whilst lacking DNA methylation functions, still retains the ability to regulate developmental gene transcription.

In order to discern if promoter hypomethylation correlated with transcription, I integrated the ERRBS and RNA-Seq datasets. This revealed that of the 59 up-regulated genes that

overlapped between these datasets in KO forebrain only 2 were hypomethylated (Serpina1b and Usp9y). In KO cerebellum a significant ($p=1.59x10^{-20}$ by hypergeometric testing) minority of 22 (See Appendix List A10) out of the 85 up-regulated genes that overlapped in the dataset also had hypomethylated promoters (Figure 5.17). This suggests that contrary to what was seen in NPCs, promoter hypomethylation is more correlated with transcriptional activation in up-regulated genes in the KO cerebellum.

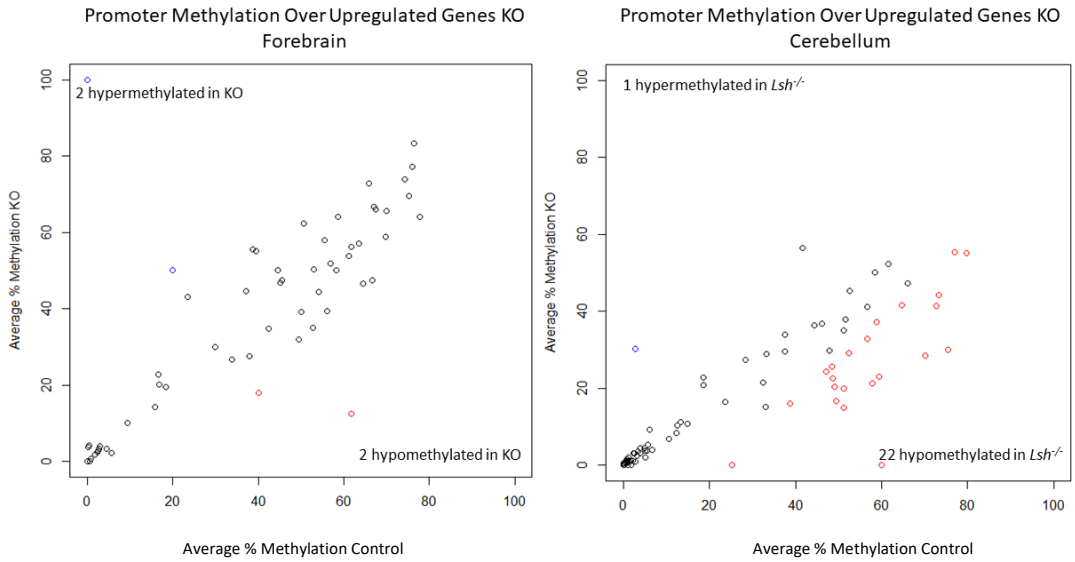


Figure 5.17 Overlap of up-regulated and hypomethylated gene promoters in KO brain. Scatter plots display the average promoter methylation of genes up-regulated in KO mouse forebrain and cerebellum. Those genes with differentially methylated promoters (defined as a change of >20%) are shown as red circles for hypomethylated and blue for hypermethylated in KO tissue.

5.2.4.3 Repeats are not enriched at up-regulated genes in KO neural tissue

As repeat transcription was shown to be up-regulated in KO neural tissue, it is possible that this would influence the expression of nearby genes. For example there may be spreading of the euchromatic state from the repetitive region to the TSS of the gene and increased availability of the transcriptional machinery recruited to the site. Another possibility is that nearby LTRs could act as alternative promoters as is well described in the literature (Dunn et al., 2003; Feuchter and Mager, 1990; Medstrand et al., 2001; Romanish, 2007). I therefore examined the repeat content of the genome surrounding the TSS of up-regulated genes using the method described in section 4.2.1.3. This revealed that promoters of up-regulated genes in KO mouse forebrain and cerebellum were actually found in genomic regions with less L1 or LTR content than non-differentially expressed genes (Figure 5.18).

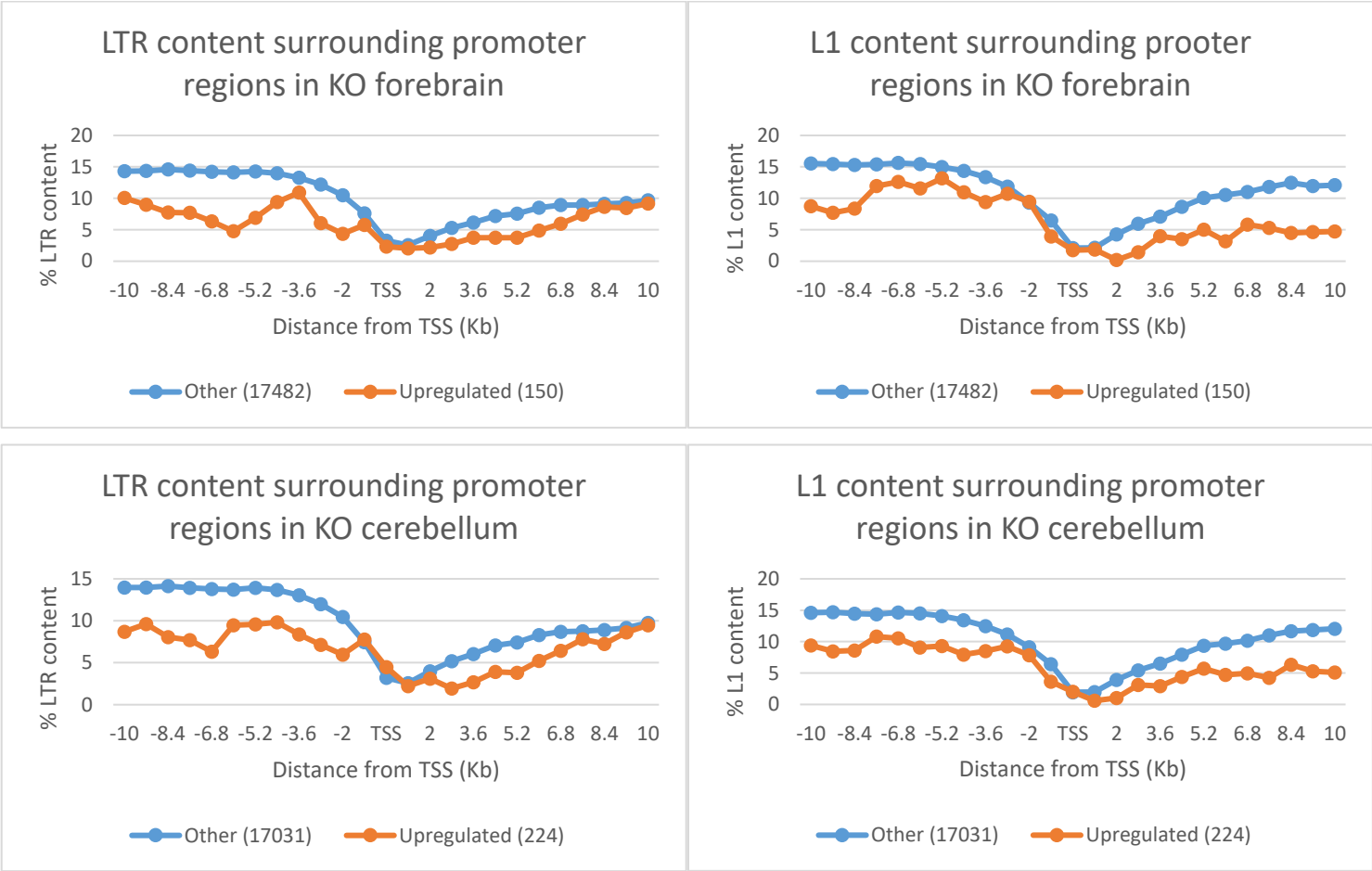


Figure 5.18. Repeat content of genome proximal to up-regulated genes in KO neural tissue. This graph divides the genome surrounding the TSS of up-regulated genes in KO forebrain and cerebellum (orange line) and genes not differentially expressed between control and KO mice (blue line) into 800bp bins 10kb up and downstream of the TSS. Each bin is represented by a point on the graph. The average percentage of base pairs within each 800bp bin that form part of an L1 or LTR annotated repeat is shown on the y axis. The number of genes in each sample is highlighted in parenthesis.

To look at the effect individual de-repressed repeats may be having on gene transcription I identified de-repressed repeats which could be considered in the regulatory domain of a gene. This was done using the default settings of the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). This tool assigns a gene basal regulatory region that extends 5kb upstream and 1 kb downstream of the TSS. The gene's regulatory domain is then extended up to the basal regulatory region of the nearest up and downstream genes (to a maximum of 1MB). This tool has been used by others to identify retrotransposons in human cells with the potential to regulate immune genes (Chuong et al., 2016). In this paper the authors elegantly demonstrated that certain LTRs in the human genome have been co-opted to act as enhancers of genes involved in the immune response. Given the gene ontology findings described in section 5.2.4.2, I was interested to investigate if de-repressed retrotransposons could be acting as enhancer / promoter elements of immune genes in the mouse model.

This analysis revealed a total of 43 genes in forebrain and 211 in cerebellum which contained up-regulated repeats in these defined regulatory regions in knockout tissue. Of these genes, 16 were found to be up-regulated in knockout cerebellum and only 1 in knockout forebrain (Appendix Table A1). Three of these up-regulated genes, namely Hamp2, Serpinb1c and Cyp2a5 were identified by the online *Interferome* tool to be interferon responsive genes. However, this is not described as their primary role and they are not described as upstream regulators of innate immune pathways. Therefore their up-regulation is not known to drive an innate immune response. This demonstrates that de-repression of retrotransposons cannot directly account for the vast majority of up-regulated genes in knockout tissue and does not explain activation of the innate immune system by proximity of immune genes to upregulated repeats. A caveat to this is that there may be up-regulated repeats missed by our analysis due to exclusion of reads that did not match our robust cut offs for unique mapping or due to polyA selection. Other potential explanations for the activation of the innate immune response are discussed in section 5.3 and chapter 6.

5.2.5 Overlap of models with human ICF syndrome

A key and exciting prospect is the potential for this knockout mouse model to be used as a tool for the investigation of neurological deficits in ICF syndrome. A note of caution is that one of the main roles of *Lsh* is methylation of repeats and there is divergence of repeat family subclases and activity between species. Therefore cross-translation of results to humans should be done with care. In saying this, models of other subtypes of ICF do show phenotypes reminiscent of the human condition (Rajshekar et al., 2018; Ueda et al., 2006).

I was interested to compare my findings with published data from human studies in ICF patients. A recent study analysed the methylome from whole blood in patients with all 4 known subtypes of ICF syndrome. Findings revealed that although there was overlap of hypomethylation at pericentromeric repeats and some common loci, methylomes could actually distinguish between subtypes. Although the majority of hypomethylation occurred at intergenic regions, there were a number of gene clusters which were found to be hypomethylated in patients, these were the PCDH cluster (hypomethylated in all subtypes but more so in DNMT3b subtypes) and the OR and KRTAP gene clusters which were hypomethylated in ICF2-4 subtypes (Velasco et al., 2018a).

I therefore analysed methylation of these gene clusters in both my cell and mouse models. This analysis revealed a trend for hypomethylation of promoters in all of these gene clusters in each of the models. Increased expression across the PCDH locus was also found (Figure 5.19). Although expression was unchanged in the other clusters these genes are not normally expressed in control cells.

These similarities between human sites of hypomethylation and findings in the mouse lend some support to the use of this mouse as a model of ICF syndrome. Ideally more comparisons would be made with neural tissue or neural cells derived from ICF type 4 patients if this becomes available. Further behavioural testing in the mouse model would also be required to demonstrate any similarities in phenotype.

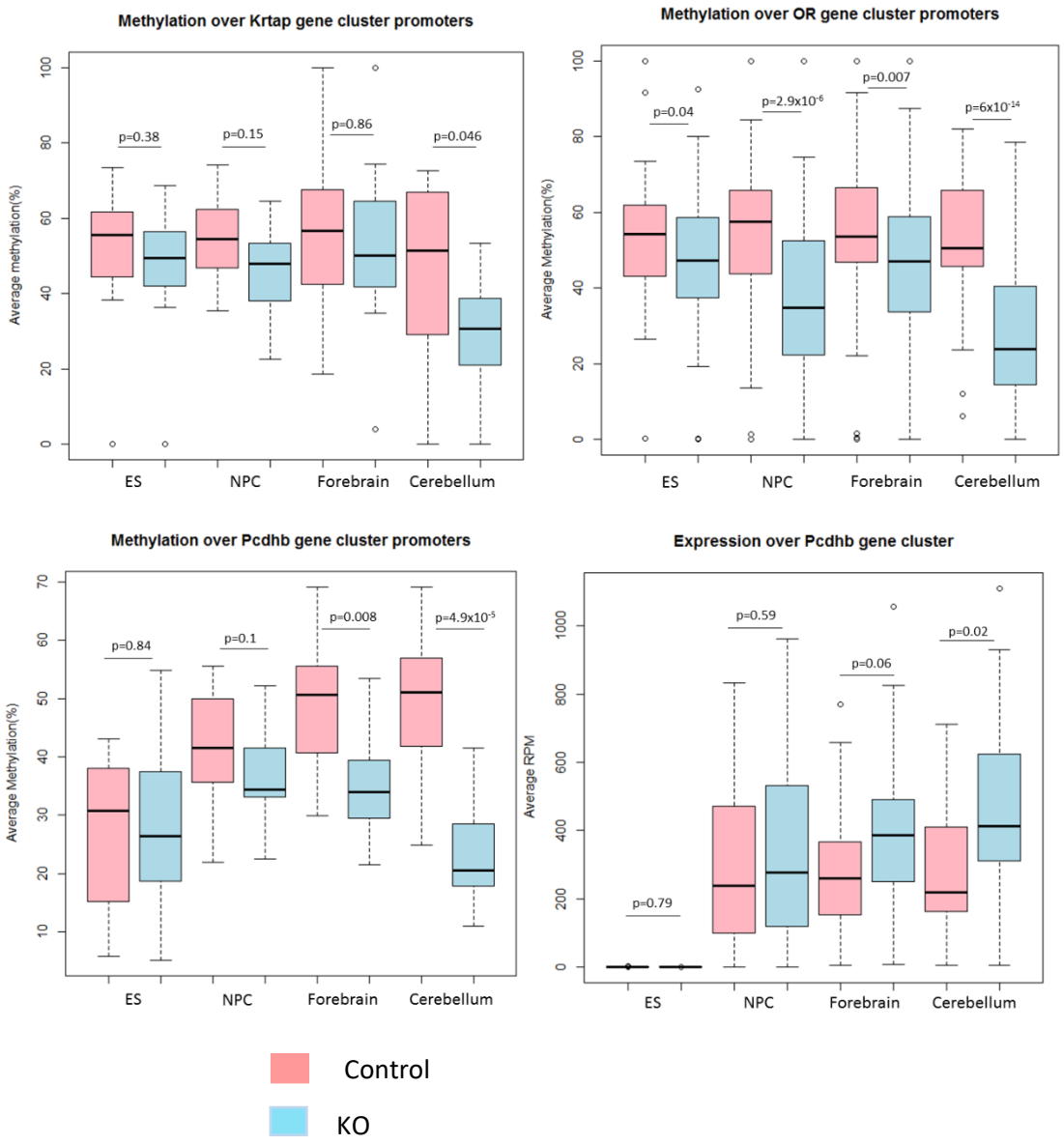


Figure 5.19. Methylation over gene clusters hypomethylated in ICF patients. Box plots display the average methylation across promoters of gene clusters; protocadherin B (*Pcdhb*), olfactory receptors (*Or*) and Keratin associated proteins (*Krtap*) in mouse neural tissue and *Lsh*$^{-/-}$ and WT ES and NP cells. The final boxplot displays the average expression across the *Pcdhb* gene cluster in mouse neural tissue and *Lsh*$^{-/-}$ and WT ES and NP cells extracted from RNA-Seq data expressed in reads per million (RPM). P values calculated from Wilcoxon rank sum test are displayed on the graph.

5.3 Discussion

In this chapter I have presented the characterisation of a novel mouse model with a targeted deletion of *Lsh* in neural tissue. It is the first mutant *Lsh* model that survives to adulthood. Furthermore it is the first model surviving to adulthood, with hypomethylated neural tissue as a result of abnormal DNA methylation across all neural cell types during development. Not only did this model permit me to investigate the requirement for *Lsh* in DNA methylation at a later stage in neurodevelopment than has previously been investigated, but it also allowed me to investigate the consequences of the absence of *Lsh* during neurodevelopment on the mature brain.

5.3.1 *Lsh* is required during development for appropriate DNA methylation and repression of repeats in the mature brain

I have demonstrated, by ERRBS, a requirement for *Lsh* during differentiation of NPCs for appropriate DNA methylation at repetitive elements, namely satellites, LTRs and LINE1s. As *Lsh* is not expressed in the mature brain, it demonstrates that an abnormal methylome deposited during neurodevelopment persists into later life. This role for *Lsh* at later stages of cell differentiation has not previously been described and is interesting given the fact that *Lsh* is down-regulated upon differentiation. Furthermore, de-repression of repetitive elements was also seen in the KO adult brain.

This hypomethylation and subsequent repeat de-repression was more pronounced in cerebellar compared to forebrain tissue. One explanation for this could be that cerebellar Purkinje cells are proposed to undergo a more extensive de-methylation, re-methylation event than neurons of the forebrain (Zhou et al., 2016). It would be interesting to investigate if this de-repression of repeats results in an increase in retrotransposition events.

Given the greater propensity for $Lsh^{-/-}$ cells to differentiate towards neural lineage and the mis-regulation of expression of genes involved in developmental processes uncovered in chapter 3, I was interested to examine the morphology of the KO mouse brain. One might expect this to be perturbed given the requirement for strict control of spatiotemporal gene expression during neurogenesis. The KO model, however, demonstrated that despite the gross changes in the neural methylome that occur in the absence of *Lsh*, neurodevelopment proceeds unperturbed with no severe effects on mature brain morphology or behaviour. Although this model does not ascertain that this would be the case if *Lsh* is absent from the very beginning of development, there are also no reported gross morphological brain changes in full $Lsh^{-/-}$ models (Geiman et al., 2001). Cerebral malformations have been described in 5 patients with ICF syndrome to date. These include macrocephaly, cortical

atrophy and corpus callosum hypoplasia (Weemaes et al., 2013). It should be noted that the majority of ICF patients do not undergo brain imaging therefore the true incidence is unknown. A large proportion of patients with ICF suffer from intellectual disability of varying degrees. This mouse model displayed no gross abnormalities in working memory or gross motor abilities. It would be very interesting, with larger numbers of mice, to carry out more sensitive testing of learning and memory to assess if there are any behavioural phenotypes akin to ICF syndrome.

5.3.2 Immune genes are up-regulated in the *Lsh* KO mouse brain

Of particular interest is the apparent activation of the immune system in the KO mouse brain, as evidenced by up-regulation of genes involved in the innate immune response in the RNA-Seq data. This analysis suggests that this was not a direct consequence of up-regulated repeats regulating the transcription of these genes as has been shown for some individual repeats in human cells (Chuong et al., 2016). A strong possibility is that the innate immune system is responding to the inappropriate accumulation of transcripts from retrotransposons via the cytosolic nucleic acid sensing pathway.

The innate immune system is a first line defence against viral pathogens. Viral nucleic acids and proteins are recognised as non-self pathogen-associated molecular patterns (PAMPs) by host nucleic acid-binding pattern-recognition receptors (PRRs).

It is now recognised that transcripts from endogenous retroviruses can also be recognised as PAMPs by these receptors and lead to human disease. The classical example of this being the autoimmune disease, Aicardi-Goutières's Syndrome (AGS), caused by mutations in *Trex1.* This gene encodes a protein with a role in metabolizing reverse-transcribed cDNAs. In these patients, intracellular accumulation of endogenous retroelement transcripts is linked to activation of the interferon response leading to clinical symptoms of neurological dysfunction and cutaneous inflammation (Stetson et al., 2008).

Products of endogenous retroelement transcription could act as PAMPs by several different means. Firstly, the transcripts could be translated for the production of viral proteins leading to budding of viral particles which could be recognised by PAMPs on the plasma membrane. Secondly, cDNA or intermediates generated during reverse transcription could be detected by cytosolic PAMPs (Hurst and Magiorkinis, 2015).

The major families of PRRs are; the Toll-like receptors (TLRs) present on the plasma and endosomal membrane (O'Neill et al., 2013), cGAMP synthase (cGAS) which detects ssDNA, dsDNA and RNA:DNA hybrids in the cytosol (Gao et al., 2013), the DNA-dependent activator of interferon regulatory factors (DAI) which detects cytosolic dsDNA (Takaoka et al., 2007)

and the RIG-1 like receptors (RLRs) which detect dsRNA (Loo and Gale, 2011). DAI was found to be up-regulated in the KO mouse forebrain and cerebellum.

Detection of PAMPs by these PRRs activates complex signalling cascades leading to altered gene expression and ultimately production of pro-inflammatory cytokines and interferons. Activation of interferon signalling gives rise to induction of IRF and STAT TFs which bind to cis-regulatory elements upstream of interferon responsive genes (Crosse et al., 2018). Some interferon stimulated genes have been shown to directly inhibit the viral life cycle and enhance early innate immune signalling. Examples of these include DDX60 which has the ability to degrade viral RNA as well as augmenting the RLR pathway (Oshiumi et al., 2015). A number of members of the TRIM family also have the ability to directly restrict viruses and also enhance PAMP detection of viral nucleic acids (Carthagena et al., 2009). PKR is another protein which can detect dsRNA in the cytoplasm and augment other innate RNA detection pathways (Der et al., 1997; Pham et al., 2016) (Figure 5.20).

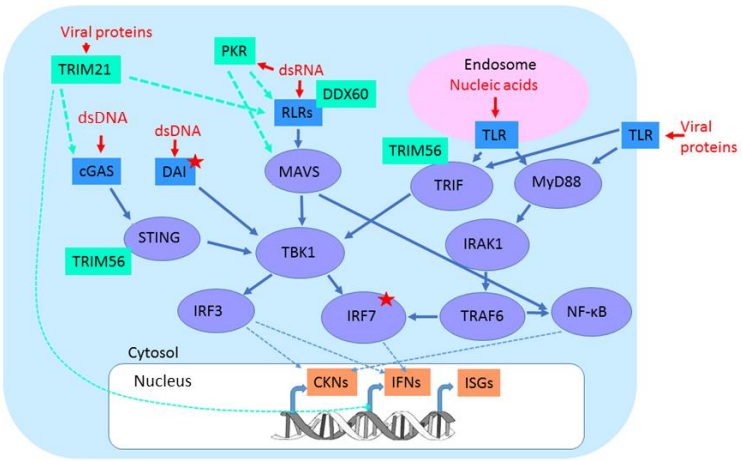


Figure 5.20. Innate immune signalling. This illustration provides a simplified view of innate immune signalling cascades in response to detection of viral / endogenous retroviral proteins and nucleic acids. See text for full description. Host nucleic acid-binding pattern-recognition receptors (PRRs), highlighted in blue boxes detect pathogen-associated molecular patterns (PAMPs), highlighted in red. This leads to the induction of complex signalling cascades, highlighted in purple which lead to the production of proinflammatory cytokines (CKNs) and interferons (IFNs) and downstream activation of interferon stimulated genes (ISGs). A number of these interferon responsive genes have been shown to play direct roles in detection and inhibition of viruses as well as augmenting the innate immune response (highlighted in green). Red stars highlight genes found in up-regulated gene lists from KO mouse neural tissue. (Figure adapted from (Crosse et al., 2018))

Alternatively, up-regulation of immune genes could reflect a Senescence-associated secretory phenotype (SASP) secondary to another cause such as DNA damage. However, I

did not find any induction of p53 (a TF with key roles in modulating cellular senescence) or related gene targets (Appendix Table A2). This will be discussed further in chapter 6.

These findings open up further exciting avenues of research into the aetiology of intellectual disability in ICF.

Chapter 6. Discussion

6.1 Summary

A number of neurodevelopmental disorders are linked with deficits in the DNA methylation pathway. There has also been a recent upsurge in the amount of research investigating altered methylomes in health and disease, particularly in my own field of psychiatry (Liu et al., 2018). Despite this, we still do not fully understand the mechanisms underlying DNA methylation reprogramming during neurodevelopment. The aim of my thesis was to broaden our understanding of this process by investigating the role of a particular protein, *Lsh,* and its contribution to DNA methylation at this time.

How defects in DNA methylation pathways during development may affect the mature brain and lead to neurological symptoms is unknown. This is largely a result of a lack of suitable models. *Lsh* and *Dnmt* mouse knockouts are lethal (Geiman et al., 2001; Okano et al., 1999; Walsh and Bestor, 1999). Even neural targeted *Dnmt* knockouts are limited in their use as models for investigating this question due to: lethality, lack of effects on methylation in post-mitotic neurons or only a subpopulation of neurons being targeted (Fan et al., 2001; Feng et al., 2010a; Golshani et al., 2005; Hutnick et al., 2009; Nguyen et al., 2007).

At the beginning of this project little was known about *Lsh*. With the recent finding of a mutation in this gene being causative for ICF syndrome (Thijssen et al., 2015), more has been published on this subject. *Lsh* is down-regulated upon differentiation of mESCs and during development, however most studies have focused their examination on end-state somatic tissue rather than during the differentiation process. Furthermore, the majority of research on this protein has focused on specific loci rather than looking at genome-wide consequences of its loss and has not fully investigated the consequences of *Lsh* inactivation on transcription states.

The aim of this project was to investigate the potential contribution of *Lsh* to DNA methylation during neurogenesis. This was done using a CRISPR generated *Lsh*$^{-/-}$ mES cell line and inducing neural differentiation using an N2B27 protocol. This allowed assessment of its role in DNA methylation at an early stage of neurogenesis, i.e. formation of neural progenitor cells. In contrast to the majority of previous studies, a genome-wide approach was adopted, utilising ERRBS and RNA-Seq to assess the methylome and transcriptome respectively. The discovery of the misregulation of developmental genes in *Lsh*$^{-/-}$ cells led onto a genome wide analysis of Polycomb distribution.

In the second part of this project a novel *Nestin-Cre* targeted *Lsh*$^{-/-}$ mouse generated by Dr Ian Adams Laboratory was also employed. This created a targeted deletion of *Lsh* in neural

progenitor cells. In contrast to the previous models mentioned, this mouse with inactivation of exon 12 of *Lsh* in neural lineage cells: survived to adulthood, demonstrated hypomethylated neural tissue and deletion was not targeted to only a subpopulation of neurons. As *Lsh* was not deleted until NPC stage (beginning E10 when *Nestin* is expressed), the contribution of *Lsh* to DNA methylation at a later stage of neurodevelopment could be examined. This was done through genome-wide interrogation of the methylome and transcriptome of forebrain and cerebellar tissues.

It was hoped that by using this model, we would gain a better understanding of the consequences of aberrant DNA methylation during development on the mature brain with the possibility of gaining new insights into the aetiology of the neurological symptoms found in ICF syndrome.

The major findings of this body of work have resulted in the following conclusions:

- *Lsh* contributes to DNA methylation of repetitive elements at both early and later stages of neurogenesis
- *Lsh* is required for appropriate distribution of Polycomb
- *Lsh* plays a role in regulating Polycomb targeted developmental gene transcription which may be independent of its role in DNA methylation
- *Lsh* is required to repress repeat transcription in the mature mouse brain
- Innate immune pathways are activated in the murine brain in the absence of *Lsh*
- The role of *Lsh* is context dependent

6.2 *Lsh* contributes to DNA methylation during neurogenesis

The requirement for *Lsh* for appropriate global DNA methylation during early neurogenesis was demonstrated by LC-MS. This revealed no difference in global 5mC between WT and *Lsh*$^{-/-}$ ES cells, but following differentiation to NPCs the knockout cells became hypomethylated, reaching only ~70% of the methylation level of WT cells. ERRBS data revealed a specific loss of methylation at repeat sequences following differentiation and at a number of single copy gene promoters.

The established role for *Lsh* is in *de novo* methylation. This was concluded largely as a result of work transfecting episomal vectors into *Lsh*$^{-/-}$ MEFs. These experiments revealed a requirement for *Lsh* to establish novel methylation patterns, but no requirement for maintenance of previously methylated vectors (Zhu et al., 2006). Furthermore, the association of *Dnmt3b* with repeat sequences in ES cells is impaired in the absence of *Lsh* (Ren et al., 2015). Analysis of my ERRBS data revealed that at L1 repeats, methylation was

gained in both WT and *Lsh*$^{-/-}$ cells upon differentiation to NPCs, but to a lesser extent in the absence of *Lsh*. This would be in keeping with the aforementioned role in *de novo* methylation. This replicates the pattern seen in another study examining L1 repeat methylation by bisulphite sequencing following neural differentiation of *Lsh*$^{-/-}$ cells (Ren et al., 2015). Of particular interest, however, was the pattern of methylation at Satellite and LTR repeats. ERRBS revealed that whilst WT cells maintained their methylation level at these sites upon differentiation, *Lsh*$^{-/-}$ cells failed to do so. This would suggest a role for *Lsh* in maintenance methylation in this context. The reduction in global methylation upon differentiation of *Lsh*$^{-/-}$ ESCs to NPCs as demonstrated by LC-MS supports this. This could be tested by maintaining a population of dividing NPCs and repeating methylation analysis at different passages.

An alternative explanantion is that there may be further demethylation / remethylation events at these sites during the transition from ESC to NPC. Repeating methylation analysis at these sites at multiple time-points during this developmental window would be required to confirm this. Support for a role in maintenance methylation comes from the finding that knockdown of *Lsh* in primary human fibroblasts leads to progressive loss of methylation at repeat sequences (Suzuki et al., 2008), and that *Lsh* has also been shown to directly interact with the maintenance methyltransferase *Dnmt1* (Dunican et al., 2015). It would be interesting to carry out a ChIP for *Dnmt1* upon differentiation of WT and *Lsh*$^{-/-}$ cells to assess its association at Satellite and LTR sites.

Not only did the work presented in this thesis demonstrate a requirement for *Lsh* in DNA methylation from ESC to NPC differentiation, but it also revealed a previously unknown requirement for *Lsh* at later stages of neurogenesis also. This was demonstrated by methylation analysis of neural tissue from the *Nestin-Cre Lsh*$^{-/-}$ mouse, where global methylation levels in forebrain and cerebellum reached 88% and 77% of control levels respectively. In this model, *Lsh* is not depleted until NPC stage when *Nestin* is expressed. Therefore, despite the down-regulation of *Lsh* expression upon differentiation (which was confirmed in my RNA-Seq data set (Appendix Figure A4)), it is still required at these later stages for appropriate DNA methylation.

This is the first study looking at genome-wide promoter methylation levels in *Lsh*$^{-/-}$ NPCs and mouse neural tissue. This was done given the reports of selected single copy loci hypomethylation in other *Lsh*$^{-/-}$ models (Fan et al., 2005b; Xi et al., 2009; Xi et al., 2007). ERRBS data revealed hypomethylation of 395 gene promoters in *Lsh*$^{-/-}$ NPCs, 164 in KO forebrain and 708 in KO cerebellum. Given the role for Lsh in methylating repetitive elements, I considered the possibility that these promoters were hypomethylated due to their proximity to repeat sequences rather than by direct targeting by *Lsh*. Analysis revealed that

LINE1s, and to a lesser extent LTRs, were enriched in the 20kb region surrounding these hypomethylated promoter TSSs when compared to non-differentially methylated promoters in KO NPCs, forebrain and cerebellum suggesting that this was, indeed, the case.

Despite the finding that a number of promoters displayed hypomethylation in the knockout models, this did not translate into an increased transcription of these genes in NPCs or mouse forebrain. There was, however, a stronger correlation between expression and promoter methylation status in the KO cerebellum.

In this study, ERRBS was employed as the main method by which to study DNA methylation. Its many advantages include, the ability to carry out genome-wide assessment of the methylome with greater coverage than traditional RRBS methods, at reduced cost and sequencing requirements of whole-genome bisulphite sequencing. However, it does not possess the coverage of the latter, favouring CpG dense genomic regions, therefore may have missed regions where *Lsh* plays a role in methylation. This is particularly important to highlight given the recently published paper comparing the methylomes of patients with all 4 subtypes of ICF syndrome (although this only included one patient with a mutation in *Hells*). In this paper 36% of hypomethylated probes in Type 1 patients were in CGIs compared to only 4% in Type 4 where 71% were in “open sea”, defined as isolated CpGs in the rest of the genome (Velasco et al., 2018b). Additionally, sites of non-CpG methylation, which are enriched in neural tissue, will be missed. Furthermore, bisulphite sequencing cannot distinguish between 5mC and 5hmC (Nestor et al., 2010). This may be particularly relevant in these model systems given the abundance of 5hmC in neural tissue (Khare et al., 2012; Spiers et al., 2017). In order to distinguish between these modifications, methods such as methylated and 5-hydroxymethylcytosine DNA immunoprecipitation would have to be used (Lentini et al., 2018).

6.3 *Lsh* represses repeat transcription

An important consequence of this hypomethylation was the reactivation of normally repressed repetitive sequences. RNA-Seq demonstrated a derepression of LTRs and LINE1s in KO forebrain and cerebellum.

PolyA-selected RNA-Seq was employed as opposed to total RNA-Seq. This has the advantage of selecting against ribosomal RNA, which can overwhelm the data by contributing a high proportion of reads. However, recent papers have demonstrated the inferiority of polyA-selection compared to total RNA-Seq when looking at repeat sequences (Solovyov et al., 2018). This is particularly true for satellite sequences, which may explain why I found a derepression of satellites by qRT-PCR in KO cerebellum but not in the RNA-

Seq data. It would therefore be interesting to examine repeat transcription profiles using total RNA-Seq.

As discussed in chapter 1, repetitive elements are normally strictly repressed as their activity could have a broad range of detrimental consequences (Crichton et al., 2014; Garcia-Perez et al., 2016). The consequences of retrotransposition can be severe resulting in genomic instability and insertional mutagenesis. De-repression of repeat elements could have effects on gene expression independent of retrotransposition. They can also promote aberrant transcription as some L1 and LTRs contain bidirectional promoters and enhancer elements thereby promoting transcription of flanking sequences (Cruickshanks and Tufarelli, 2009; Speek, 2001). Intronic insertions can lead to aberrant splicing or premature polyadenylation. Futhermore the retrotransposon may attract repressive epigenetic modifications silencing nearby genes by spreading of heterochromatin (Cruickshanks et al., 2013). Finally, protein products of retrogenes could alter cell functions. It would be very interesting to determine if retrotransposition was occurring within the *Lsh*$^{-/-}$ models for example by whole genome sequencing or by using transposition reporters (Macia et al., 2017).

To determine if de-repressed repeat elements have effects on gene expression in my mouse model I examined the repeat content of the genome 10kb up and downstream of the TSS of those genes found to be up-regulated in the KO mouse. This revealed that the genome surrounding up-regulated genes was actually less rich in repeats than that of surrounding non-differentially expressed genes. This however does not exclude a role for individual repeats in influencing transcription. I therefore looked into this by examining the presence of up-regulated repeats within regulatory domains of up-regulated genes. This revealed that of the 150 genes up-regulated in KO forebrain only 1 had an up-regulated repeat in its regulatory domain. Of the 224 genes up-regulated in cerebellum, 16 had up-regulated repeats within their regulatory domains. Therefore the majority of differentially expressed genes do not appear to be so as a direct result of transcriptional regulatory control by up-regulated repeats. This analysis is of course limited to those repeats meeting our robust cut offs for inclusion within the analysis. Furthermore I cannot absolutely exclude that up-regulated repeats may be acting as distal enhancers given the method in which I defined regulatory domains. ChIP for enhancer associated marks such as H3K27 acetylation would have to be carried out to examine this possibility.

6.4 *Lsh* regulates the expression of developmental genes

Analysis of the transcriptome of *Lsh*$^{-/-}$ ESCs and NPCs by RNA-Seq revealed a number of differentially expressed genes (597 and 889 respectively). Ontology analysis brought to light

an enrichment of genes to do with developmental processes. As well as by DNA methylation, developmental genes are known to be regulated by the Polycomb system during differentiation. Given the lack of correlation between differential expression and promoter methylation status, it led me to consider whether mis-regulation of the Polycomb system may account for the transcriptional changes seen.

This hypothesis was drawn on our knowledge of Polycomb redistribution in other hypomethylated models, including cancer, where it is proposed that hypomethylated regions provide new binding sites for Polycomb, diluting its deposition across the genome (Dunican et al., 2013; Reddington et al., 2013). In order to investigate this, I carried out ChIP-Seq for the Polycomb repressive mark H3K27me3. This revealed a widespread depletion of this mark across promoter regions in *Lsh*$^{-/-}$ NPCs and redistribution to hypomethylated repeats suggesting a requirement for appropriate DNA methylation to allow correct targeting of Polycomb.

This was also correlated with transcriptional status, as differentially expressed genes were shown to have a greater reduction of this mark across promoter regions compared to non-differentially expressed genes. Furthermore, genes differentially expressed in *Lsh*$^{-/-}$ NPCs were enriched for H3K27me3 in WT cells compared to those that were not differentially expressed. This establishes a role for *Lsh* in regulating expression of Polycomb targeted genes and concurs with previously reported findings of aberrant *Hox* gene expression in *Lsh*$^{-/-}$ mice (Xi et al., 2007).

Initially I considered this transcriptional regulation by *Lsh* occured secondary to its effects on DNA methylation. In this model, in the absence of Lsh, repetitive elements become hypomethylated upon differentiation resulting in redistribution of Polycomb away from target sites of genes associated with development to hypomethylated repeats resulting in deregulation of developmental genes and the greater propensity for *Lsh*$^{-/-}$ cells to differentiate to neural lineage. However, I found that repression of *Hox* gene expression was rescued by the reintroduction of a mutant version of *Lsh* that did not rescue the methylation defect. This would argue against this theory. In keeping with this, *Lsh*$^{-/-(+\ mutant\ Lsh)}$ cells behaved like WT upon differentiation. Furthermore, a depletion in Polycomb was not only associated with up-regulated but also with down-regulated genes in *Lsh*$^{-/-}$ NPCs.

This therefore provides preliminary evidence for a role for Lsh in regulating developmental gene expression independent of its role in DNA methylation. This may be through direct targeting of Polycomb, indeed, *Lsh* has been found to co-immunoprecipitate with components of the PRC1 complex (Xi et al., 2007) or association with other epigenetic modifiers. For example the N-terminal portion of *Lsh* containing the coiled coil domain alone, a region often involved in protein-protein interactions, has been found to be sufficient to

silence a reporter gene. Silencing of this reporter was compromised in the presence of a deacetylase inhibitor suggesting it may achieve this suppression through interaction with HDACs (Myant and Stancheva, 2008).

There are unquestionably caveats to the above. Firstly I have used *Hox* gene expression as a proxy measure for developmental gene regulation as a whole. Furthermore, although I demonstrated hypomethylation at major satellite repeats in the $Lsh^{-/-(+mutant\ Lsh)}$ NPCs this does not necessarily indicate global hypomethylation. In the future it will be necessary to carry out H3K27me3 ChIP and RNA-Seq experiments on both the $Lsh^{-/-(+WT\ Lsh)}$ and $Lsh^{-/-(+mutant\ Lsh)}$ cells. It would also be interesting to carry out ChIP for the LSH protein itself to determine if it is in fact associated with these developmental genes during differentiation.

In the future, it will be useful to test an *Lsh* mutant spectrum for their ability to rescue the different aspects of *Lsh* function that I have identified in my work.

6.5 The role of *Lsh* is context dependent

The different models used in this thesis reveal that the role of *Lsh* is dependent upon developmental, cellular and genomic context.

Despite its high expression in ES cells, it does not appear to be playing a major role in DNA methylation at this stage as only minor degrees of hypomethylation were seen at repetitive sites despite multiple passages. $Dnmt3b^{-/-}$ / $Dnmt3a^{-/-}$ double mutant ES cells show loss of methylation after repeated passaging (>30) suggesting these methyltransferases may play a role in "filling in" hemi-methylated sites missed by *Dnmt1* (Chen et al., 2003). Therefore, it may be that *Lsh* plays some role in this. It would be interesting to see if further hypomethylation occurred following multiple passages.

As *Lsh* is not playing a major role in DNA methylation at this stage it does make one wonder what its role is, given it is so highly expressed (Assou et al., 2007). It may be that its methylation-independent function at this stage is related to the high rates of proliferation seen in these cells, given its expression in adult tissues is restricted to highly proliferative tissues. It has also been proposed to play a role in "priming" sites for DNA methylation upon differentiation by altering chromatin accessibility (Ren et al., 2018).

Certainly loss of *Lsh* did seem to affect ES cells to a certain extent in terms of transcription, as I found 597 differentially expressed genes at this stage. Gene ontology revealed an enrichment of developmental terms. At this stage there was very little alteration of H3K27me3 distribution, thereby lending further support to the hypothesis that *Lsh* may have

a more direct role, perhaps directly by chromatin remodelling or association with other epigentic modifiers, in regulating expression of developmentally important genes.

The effects of *Lsh* loss also depended upon cellular context. For example, hypomethylation and subsequent repeat de-repression occurred to a greater extent in the KO cerebellum than forebrain. This may be explained by the differences in cellular composition between these brain regions. Zhou *et al* proposed that neural cells undergo a cellular de-methylation and re-methylation (CDR) event, previously only thought to occur at the very early stages of development. The timing and extent to which this happens appears individual to each cell type and for the most part is on a small scale. The exception to this are cerebellar Purkinje cells, which undergo an extensive CDR at P14-30 in the mouse brain coinciding with dendritogenesis and synaptogenesis (Zhou et al., 2016).

Interestingly, despite gross hypomethylation, repetitive elements were not de-repressed in *Lsh*$^{-/-}$ NPCs. There could be several explanations for this. It could be that hypomethylation alone is not sufficient for their de-repression and that, for example transcription factors only expressed in mature neurons drive their expression. Alternatively there could be other factors present in NPCs repressing their transcription that are no longer present upon differentiation. Another possibility is that there may be a retrotransposition window upon NPC differentiation with the increased number of transcripts actually being secondary to an increased copy number. Finally the redistribution of Polycomb may be sufficient to repress these repeats in NPCs.

6.6 The *Nestin-Cre* targeted *Lsh* knockout mouse provides a novel model for investigating the consequences of aberrant developmental DNA methylation processes on the mature brain

Aberrations in DNA methylation deposition during development, or mutations in genes which form part of the DNA methylation machinery, are associated with neurodevelopmental disorders (Bell et al., 1991; Hendrich and Bickmore, 2001; Klein et al., 2011; Neul et al., 2010; Winkelmann et al., 2012). As well as these identified disorders with known causative mutations, there is now evidence of an altered DNA methylome in screening of those patients with no identified cause of neurodevelopmental disorder (Barbosa et al., 2018).

Understanding these processes and their impact on the mature brain is therefore important, but as previously mentioned, there is a lack of suitable models for investigating this. The new mouse model I have described in this thesis could therefore prove to be a useful tool, as it survives, with apparent health, to adulthood with hypomethylated neural tissue. The hypomethylation seen in the neural tissue of this model demonstrates that the role of a protein early in development can have a lasting impact on the neural methylome in the adult, despite the fact it is no longer present in this tissue at this stage. This allows one to study purely the effects of hypomethylation during development, as other proteins such as the *Dnmts* and *Mecp2* continue to have roles in the adult brain (Bayraktar and Kreutz, 2018; Du et al., 2016; McGraw et al., 2011; Oliveira et al., 2012).

The absence of morphological abnormalities in the knockout mouse brain, although expected as previous *Lsh*$^{-/-}$ knockout mouse models showed no gross changes in brain morphology, is intriguing given the vast disruption of DNA methylation, thought to be a key process necessary for development.

I discussed earlier the role *Lsh* plays in regulating developmental genes. One might therefore have predicted this would influence neurodevelopment, where the timing of each stage of neurogenesis is tightly controlled, and disruption of this process could result in altered cortical thickness and layering. This, however, was not found in the knockout mouse. It is also true, however, that ontology analysis did not reveal aberrant expression of developmental genes in knockout neural tissue. One could argue that this is because of the timing of *Lsh* deletion, such that by NPC stage it may have already performed its important roles in regulating transcription at these sites. It could also be that our mouse model has a truncated *Lsh* protein capable of carrying out transcriptional regulation of developmental genes whilst lacking the ability to ensure appropriate DNA methylation. This is similar to what

was seen in the *Lsh*$^{-/-(+\ mutant\ Lsh)}$ cell line which behaved more like WT cells upon differentiation. However, *Hox* genes are aberrantly expressed in previous non-targeted *Lsh* knockout mice, in which there is no reported abnormality in cortical thickness *etc.* suggesting that this has little consequence for brain development (Geiman et al., 2001).

Despite gross hypomethylation, this knockout mouse displayed no obvious behavioural phenotypes. Given the limited number of mice available, more sensitive behavioural testing for detection of learning and memory defects was not possible. It would be of great interest to carry this out in the future to determine if these mice have a phenotype comparable to the intellectual disability seen in patients with ICF syndrome.

This leads me on to consider if the mouse would be a suitable neurological model for ICF syndrome. The important caveats to take into account are that in this model *Lsh* is absent only from NPC stage whereas in patients it will be absent from the very beginning of development. Furthermore, there are distinct differences between repeat families and activity between mice and humans. In mouse, centromeric and pericentromeric regions are composed of minor and major satellites whereas in humans they are represented by alpha satellite DNA. In saying this, there is a striking resemblance of an ICF phenotype in mice carrying human ICF mutations in *Dmnt3b* (Ueda et al., 2006). The characteristic hypomethylation at satellite sequences, and the overlap between hypomethylation at particular gene clusters in our mouse model and the methylome of ICF patients, as shown in chapter 5, would lend support for the use of this mouse as a tool to investigate the pathogenesis involved in the neurological deficits seen in ICF syndrome as discussed below.

6.7 The innate immune system is activated in the absence of *Lsh*

Of particular interest was the finding that genes involved in innate immunity were up-regulated in the knockout mouse brain. A chronic inflammatory state is a common feature of many neurological disorders including MS, lupus, neurodegenerative and psychiatric conditions and therefore could be an aetiological factor in the neurological defects seen in ICF syndrome (Amor et al., 2010; Leighton et al., 2017; Najjar et al., 2013; Wee Yong, 2010). So what could be causing this inflammatory state?

A strong possibility is the activation of repeats. Innate immune detection of viral nucleic acids and proteins is achieved through pattern recognition receptors located on cell membranes and in the cytoplasm and activation of these pathways results in an interferon response (Barbalat et al., 2011). These same innate immune sensors can also detect

products of endogenous retroelement transcription. The discovery of this occurred during research into the autoimmune disease Aicardi-Goutières Syndrome (AGS) in which patients suffer from neurological dysfunction and cutaneous inflammation. This disease is caused by mutations in *Trex1.* Generation of *Trex1* deficient mice revealed activation of the interferon response and accumulation of intracellular DNAs mapping to endogenous retroelements. *Trex1* was found to play a role in metabolizing reverse-transcribed cDNAs (Stetson et al., 2008). Activation of the innate immune response by aberrant LINE1 transcription has also recently been shown to occur in ageing cells suggesting this could be a common pathway in many pathological states (De Cecco et al., 2019).

In support of this, *Dai* a key cytosolic dsDNA sensor was up-regulated in KO mouse neural tissue. Furthermore, a new zebrafish model of ICF, null for *Zbtb24,* demonstrates de-repression of pericentromeric transcripts and activation of an innate immune response that was blocked by mutating the dsRNA-sensing machinery (Rajshekar et al., 2018).

It would be interesting to test this theory in our model further by, for example, crossing the mice with *Dai*$^{-/-}$ mice or treating with reverse transcriptase inhibitors in conjunction with behavioural testing. If this indeed does play a role in the pathogenesis of ICF it does offer exciting therapeutic options. For example reverse transcriptase inhibitors already used in HIV and currently under trial in AGS may be of benefit (https://clinicaltrials.gov/ct2/show/NCT02363452).

Hypomethylation of repeats could cause the observed up-regulation of immune genes by alternative mechanisms. Firstly, hypomethylated gene promoters in the *Lsh*$^{-/-}$ models, when subjected to ontology analysis, were found to be enriched for innate immune terms and were enriched for LTR and LINE1 repeats in the 20kb region surrounding their TSS. It could therefore be that hypomethylation of these gene promoters, and subsequent activation is simply due to their proximity to hypomethylated repeats. However, *Lsh*$^{-/-}$ NPCs displayed hypomethylation at immune response genes but no activation of the immune response pathway. Another explanation could be the use of hypomethylated LTRs as promoters or enhancers of immune genes. Specific examples of this have been identified by the Feschotte group in human cells who demonstrated that many LTRs contained binding sites for interferon-inducible transcription factors, that these elements were enriched near genes annotated with immune functions, and gained H3K27ac (a hallmark of enhancer activity) on treatment with interferons. Deletion of one particular ERV resulted in failure of activation of the nearby immune gene and a subsequent deficit in the inflammatory response, whereas this ERV was sufficient to drive interferon inducible reporter expression (Chuong et al., 2016). However, as previously mentioned, I examined the presence of up-regulated repeats in the regulatory domains of up-regulated genes. Of the 17 up-regulated genes across neural

tissue found to have an up-regulated repeat within their regulatory domain, 3 were interferon responsive genes although are not upstream regulators of immune response pathways. Therefore this explanation for immune activation appears less likely in my model.

It would also be very interesting to investigate if there are increased rates of retrotransposition within the knockout mouse brain. This may lead to “sick” neurons if such events resulted in insertional mutagenesis, double strand breaks or genomic rearrangements. These dysfunctional neurons may then result in an inflammatory response. In the brain, somatic retrotransposition has been shown to occur and possibly occurs even in postmitotic neurons (Coufal et al., 2009; Kuwabara et al., 2009; Macia et al., 2017; Muotri et al., 2005). However we lack understanding as to how, or why, this occurs and its influence on neural biology and disease. Having a model in which there are increased rates of retrotransposition would therefore prove a very useful tool for investigation of this process.

There are other potential reasons why there may be an up-regulation of interferon stimulated genes. It is reminiscent of a senescence-associated secretory (SASP) phenotype. Senescence was a term originally coined to describe an irreversible loss of proliferative potential in cells capable of self-renewal as a consequence of DNA damage. The term has recently, however, been applied to post-mitotic neurons which can respond to DNA damage using similar pathways to mitotic cells (Fielder et al., 2017). In this model there may be activation of the DNA damage response pathway either through increased rates of damage or ineffective repair. As *Lsh* is not expressed in adult tissue, ineffective repair is unlikely. DNA damage could be incurred during retrotransposition events as described above or it could be that hypomethylated repeats results in genomic instability even in postmitotic cells. However, RNA-Seq data did not show evidence for a DNA damage response as there was no differential expression of p53 target genes (as defined by Wei *et al* (Wei et al., 2006)) in mouse neural tissue (Appendix Table A2). Interrogation of the knockout mouse brains by gamma-H2AX and TUNEL staining to detect any evidence of DNA double strand breaks and apoptosis would be desirable to confirm this. A recent study by Han *et al*, revealed a small increase of 3% in apoptotic cells in the subventricular zone in *Lsh*$^{-/-}$ embryonic brain (Han et al., 2017). However, the lack of any difference in cortical thickness or brain weight in my model would argue against a large-scale apoptosis of neurons in our model.

6.8 Final comments

In this thesis, I have demonstrated a role for *Lsh* in DNA methylation throughout neurogenesis. This role in DNA methylation was dependent upon the helicase domain as

demonstrated by the failure of reintroduction of *Lsh* carrying a mutation in this domain to rescue the methylation defect in knockout cells. As the helicase domain has key roles in chromatin remodelling, this further supports the theory that Lsh is required for nulceosomal remodelling at repeat regions to allow access for the DNA methyltransferases.

Furthermore, this work has uncovered a role for *Lsh* in regulating the expression of polycomb targeted developmental genes. In concurrence with other hypomethylated models, lack of *Lsh* resulted in a redistribution of the repressive Polycomb mark H3K27me3. Intriguingly, however, I discovered preliminary evidence to suggest that *Lsh* may regulate developmental gene expression independently of its role in DNA methylation or chromatin remodelling. I drew these conclusions from the fact that *Lsh* carrying a mutation in the helicase domain which did not rescue the methylation defect did rescue Hox gene repression. It is therefore possible that *Lsh* interacts with epigenetic regulators, for example via the coiled-coil or ATP domain, to control transcription at these sites (Figure 6.1).

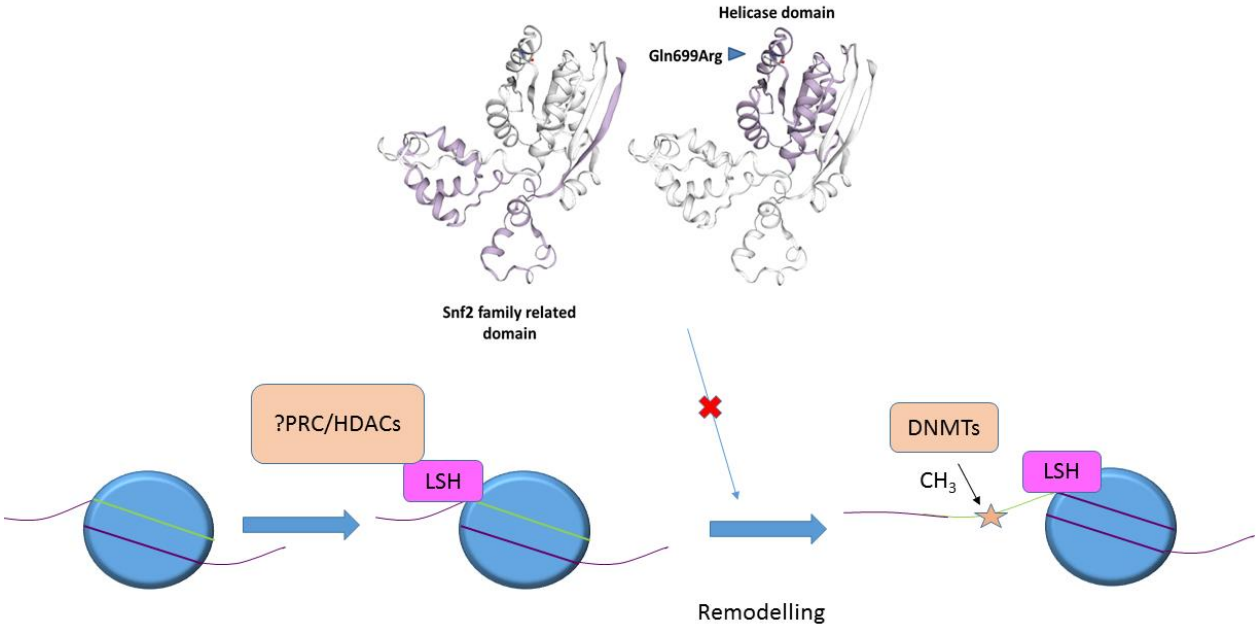


Figure 6.1. Lsh methylation model. This body of work demonstrates the necessity of the helicase domain of *Lsh* for its roles in DNA methylation. Our model is therefore that chromatin remodelling allows access for the DNA methylation machinery to repetitive sites. In contrast this domain is not required for the regulation of developmental gene expression. Therefore we propose that this role is achieved through *Lsh* interacting with other epigenetic regulators such as the Polycomb complexes or HDACs at its coiled coil or Snf2 ATP domain which is intact in our *Lsh*$^{-/-(+mutant\ Lsh)}$ system.

In the second part of this thesis I described the investigation of a novel, neurally targeted, *Lsh*$^{-/-}$ mouse which demonstrates hypomethylation of repeats resulting in their transcriptional derepression. Of particular interest was the finding of activation of the innate immune system in KO neural tissue. This led me to propose a model by which aberrant repeat transcription

results in an immune response due to cellular detection of cytoplasmic nucleic acids generated via reverse transcription. This opens up exciting new avenues of research into the underlying pathology of the neurological deficits seen in ICF syndrome and potential therapeutic options.

References

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K.*, et al.* (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics *65*, 293-298.

Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., and Bekman, E. (2009). Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. PloS one *4*, e6286.

Achour, M., Jacq, X., Ronde, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A.D.*, et al.* (2008). The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. Oncogene *27*, 2187-2197.

Adams, I.R. (2017). Retrotransposons and the Mammalian Germline. In Human Retrotransposons in Health and Disease, G. Cristofari, ed. (Cham: Springer International Publishing), pp. 1 -28.

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L.*, et al.* (2012). Base -Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. PLoS genetics *8*.

Alghamdi, H.A., Tashkandi, S.A., Alidrissi, E.M., Aledielah, R.D., AlSaidi, K.A., Alharbi, E.S., Habazi, M.K., and Alzahrani, M.S. (2018). Three Types of Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome Identified by Whole-Exome Sequencing in Saudi Hypogammaglobulinemia Patients: Clinical, Molecular, and Cytogenetic Features. Journal of Clinical Immunology *38*, 847-853.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proceedings of the National Academy of Sciences of the United States of America *51*, 786-794.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol *19*, 229-244.

Amor, S., Puentes, F., Baker, D., and van der Valk, P. (2010). Inflammation in neurodegenerative diseases. Immunology *129*, 154-169.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E.*, et al.* (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nature cell biology.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.

Assou, S., Le Carrour, T., Tondeur, S., Strom, S., Gabelle, A., Marty, S., Nadal, L., Pantesco, V., Reme, T., Hugnot, J.P.*, et al.* (2007). A meta -analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells *25*, 961-973.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi -methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.

Baets, J., Duan, X., Wu, Y., Smith, G., Seeley, W.W., Mademan, I., McGrath, N.M., Beadell, N.C., Khoury, J., Botuyan, M.V.*, et al.* (2015). Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. Brain.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M.*, et al.* (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature *479*, 534-537.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E., and Gottlieb, D.I. (1995). Embryonic stem cells express neuronal properties in vitro. Dev Biol *168*, 342-357.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by hist one modifications. Cell Res *21*, 381-395.

Bansod, S., Kageyama, R., and Ohtsuka, T. (2017). Hes5 regulates the transition timing of neurogenesis and gliogenesis in mammalian neocortical development. Development (Cambridge, England) *144*, 3156-3167.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc ̧ his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science (New York, NY) *354*, 909-912.

Barbalat, R., Ewald, S.E., Mouchess, M.L., and Barton, G.M. (2011). Nucleic acid recognition by the innate immune system. Annu Rev Immunol *29*, 185-214.

Barber, B.A., and Rastegar, M. (2010). Epigenetic control of Hox genes during neurogenesis, development, and disease. Ann Anat *192*, 261-274.

Barbosa, M., Joshi, R.S., Garg, P., Martin-Trujillo, A., Patel, N., Jadhav, B., Watson, C.T., Gibson, W., Chetnik, K., Tessereau, C.*, et al.* (2018). Identification of rare de novo epigenetic variations in congenital disorders. Nature Communications *9*, 2064.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High -resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bayraktar, G., and Kreutz, M.R. (2018). N euronal DNA Methyltransferases: Epigenetic Mediators between Synaptic Activity and Gene Expression? The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry *24*, 171-185.

Beaujean, N., Hartshorne, G., Cavilla, J., Taylor, J., Gardner, J., Wilmut, I., Meehan, R., and Young, L. (2004a). Non -conservation of mammalian preimplantation methylation dynamics. Current biology : CB *14*, R266-267.

Beaujean, N., Taylor, J., Gardner, J., Wilmut, I., Meehan, R., and Young, L. (2004b). Effect of limited DNA methylation reprogramming in the normal sheep embryo on somatic cell nuclear transfer. Biology of reproduction *71*, 185-193.

Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U.*, et al.* (1991). Physical mapping

across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell *64*, 861-866.

Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., and Tristem, M. (2007). Rate of recombinational deletion among human endogenous retroviruses. J Virol *81*, 9437-9442.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A biva lent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res *8*, 1499-1504.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y.*, et al.* (2014). Variant PRC1 complex -dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Bodea, G.O., McKelvey, E.G.Z., and Faulkner, G.J. (2 018). Retrotransposon -induced mosaicism in the neural genome. Open Biology *8*.

Bogdanovic, O., and Lister, R. (2017). DNA methylation and the preservation of cell identity. Curr Opin Genet Dev *46*, 9-14.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature genetics *42*, 1093-1100.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science (New York, NY) *317*, 1760-1764.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science (New York, NY) *294*, 2536-2539.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S.*, et al.* (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nature cell biology *6*, 97-105.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ -line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255-256.

Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A., and Cedar, H. (1993). The ontogeny of allele -specific methylation associated with imprinted genes in the mouse. The EMBO Journal *12*, 3669-3677.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N.*, et al.* (2012). Polycomb associates genome -wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell stem cell *10*, 157-170.

Brown, S.E., Suderman, M.J., Hallett, M., and Szyf, M. (20 08). DNA demethylation induced by the methyl-CpG-binding domain protein MBD3. Gene *420*, 99-106.

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J Biol Chem *278*, 823-828.

Burlingame, R.W., Love, W.E., Wang, B.C., Hamlin, R., Nguyen, H.X., and Moudrianakis, E.N. (1985). Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 A. Science (New York, NY) *228*, 546-553.

Burrage, J., Termanis, A., Geissner, A., Myant, K., Gordon, K., and Stancheva, I. (2012). The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. J Cell Sci *125*, 5524-5534.

Cai, C., and Grabel, L. (20 07). Directing the differentiation of embryonic stem cells to neural stem cells. Dev Dyn *236*, 3255-3266.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, NY) *298*.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F., and Fraser, P. (2002). Long -range chromatin regulatory interactions in vivo. Nature genetics *32*, 623-626.

Carthagena, L., Bergamaschi, A., Luna, J.M., David, A., Uchil, P.D., Margottin-Goguet, F., Mothes, W., Hazan, U., Transy, C., Pancino, G.*, et al.* (2009). Human TRIM gene expression in response to interferons. PloS one *4*, e4894.

Champagne, K.S., and Kutateladze, T.G. (2009). Structural insight into histone recognition by the ING PHD fingers. Curr Drug Targets *10*, 432-441.

Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science (New York, NY) *354*, 468-472.

Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology *24*, 9048-9058.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol *23*, 5594-5605.

Choi, J.K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol *11*, R70.

Chowdhury, B., Cho, I.-H., and Irudayaraj, J. (2017). Technical advances in global DNA methylation analysis in human cancers. Journal of Biological Engineering *11*, 10.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404-416.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083-1087.

Cikos, S., Bukovská, A., and Koppel, J. ⁀2007⁀. Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. BMC Mol Biol *8*, 113.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem *280*, 5563-5570.

Cooper, D.N., and Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Human genetics *83*, 181-188.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R.*, et al.* (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep *7*, 1456-1470.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Emb ryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Costa, F.F., Paixao, V.A., Cavalher, F.P., Ribeiro, K.B., Cunha, I.W., Rinck, J.A., Jr., O'Hare, M., Mackay, A., Soares, F.A., Brentani, R.R.*, et al.* (2006). SATR -1 hypomethylation is a common and early event in breast cancer. Cancer Genet Cytogenet *165*, 135-143.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature *460*, 1127-1131.

Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature *274*, 775-780.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., and Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci *71*, 1581-1605.

Crosse, K.M., Monson, E.A., Beard, M.R., and Helbig, K.J. (2018). Interferon -Stimulated Genes as Enhancers of Antiviral Innate Immune Signaling. J Innate Immun *10*, 85-93.

Cruickshanks, H.A., and Tufarelli, C. (2009). Isolation of cancer -specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics *94*, 397-406.

Cruickshanks, H.A., Vafadar-Isfahani, N., Dunican, D.S., Lee, A., Sproul, D., Lund, J.N., Meehan, R.R., and Tufarelli, C. (2013). Expression of a large LINE -1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. Nucleic Acids Res *41*, 6857-6869.

De Cecco, M., Ito, T., Petrashen, A.P., Elias, A.E., Skvir, N.J., Criscione, S.W., Caligiana, A., Brocculi, G., Adney, E.M., Boeke, J.D.*, et al.* (2019). L1 drives IFN i n senescent cells and promotes age-associated inflammation. Nature *566*, 73-78.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A.*, et al.* (2011). Mutations in ZBTB2 4 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Am J Hum Genet *88*, 796-804.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. PLoS genetics *7*.

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., and Muegge, K. (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nature cell biology *8*, 1448-1454.

Deichmann, U. (2016). Epigenetics: The origins and evolution of a fashionable topic. Dev Biol *416*, 249-254.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & development *15*, 2940-2944.

Der, S.D., Yang, Y.L., Weissmann, C., and Williams, B.R. (1997). A double -stranded RNA-activated protein kinase-dependent pathway mediating stress-induced apoptosis. Proceedings of the National Academy of Sciences of the United States of America *94*, 3279-3283.

Desikan, R.S., and Barkovich, A.J. (2016). Malformations of cortical development. Annals of neurology *80*, 797-810.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem *285*, 26114-26120.

Du, F., Nguyen, M.V., Karten, A., Felice, C.A., Mandel, G., and Ballas, N. (2016). Acute and crucial requirement for MeCP2 function upon transition from early to late adult stages of brain maturation. Hum Mol Genet *25*, 1690-1702.

Dubois, N.C., Hofmann, D., Kaloulis, K., Bishop, J.M., and Trumpp, A. (2006). Nestin -Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues. Genesis *44*, 355-360.

Dunican, D.S., Cruickshanks, H.A., Suzuki, M., Semple, C.A., Davey, T., Arceci, R.J., Greally, J., Adams, I.R., and Meehan, R.R. (2013). Lsh regu lates LTR retrotransposon repression independently of Dnmt3b function. Genome Biol *14*, R146.

Dunican, D.S., Pennings, S., and Meehan, R.R. (2015). Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with Dnmt1. Biomed Res Int *2015*, 740637.

Dunn, C.A., Medstrand, P., and Mager, D.L. (2003). An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. Proceedings of the National Academy of Sciences of the United States of America *100*, 12841-12846.

Dwyer, N.D., Chen, B., Chou, S.J., Hippenmeyer, S., Nguyen, L., and Ghashghaei, H.T. (2016). Neural Stem Cells to Cerebral Cortex: Emerging Mechanisms Regulating Progenitor Behavior and Productivity. The Journal of neuroscience : the official journal of the Society for Neuroscience *36*, 11394-11401.

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clinical immunology (Orlando, Fla) *109*, 17-28.

Ehrlich, M., Jackson, K., and Weemaes, C. (2006). Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet J Rare Dis *1*, 2.

Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T., and Hanash, S.M. (2008). ICF, An Immunodeficiency Syndrome: DNA Methyltransferase 3B Involvement, Chromosome Anomalies, and Gene Dysregulation. Autoimmunity *41*, 253-271.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A*., et al.* (2012). Histone H2A mono -ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS genetics *8*, e1002774.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A*., et al.* (2012). Single -neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell *151*, 483-496.

Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P.P., Kuhn, R.*, et al.* (2001). DNA hypomethylation perturbs the f unction and survival of CNS neurons in postnatal animals. The Journal of neuroscience : the official journal of the Society for Neuroscience *21*, 788-797.

Fan, G., Martinowich, K., Chin, M.H., He, F., Fouse, S.D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H.*, et al.* (2005a). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. Development (Cambridge, England) *132*, 3345-3356.

Fan, T., Hagan, J.P., Kozlov, S.V., Stewart, C.L., and Muegge, K. (2005b). Lsh contr ols silencing of the imprinted Cdkn1c gene. Development (Cambridge, England) *132*, 635-644.

Fan, T., Schmidtmann, A., Xi, S., Briones, V., Zhu, H., Suh, H.C., Gooya, J., Keller, J.R., Xu, H., Roayaei, J.*, et al.* (2008). DNA hypomethylation caused by Lsh del etion promotes erythroleukemia development. Epigenetics *3*, 134-142.

Fan, T., Yan, Q., Huang, J., Austin, S., Cho, E., Ferris, D., and Muegge, K. (2003). Lsh - deficient murine embryonal fibroblasts show reduced proliferation with signs of abnormal mitosis. Cancer Res *63*, 4677-4683.

Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., and Magnuson, T. (1998). The Polycomb - group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development (Cambridge, England) *125*, 4495-4506.

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., and Langst, G. (2011). Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res *39*, 6956-6969.

Feng, J., Chang, H., Li, E., and Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. J Neurosci Res *79*, 734-746.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010a). Dnmt1 and Dnmt3a maintain DNA meth ylation and regulate synaptic function in adult forebrain neurons. Nature neuroscience *13*, 423-430.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E.*, et al.* (2010b). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America *107*, 8689-8694.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet *12*, 565-575.

Feuchter, A., and Mager, D. (1990). Functional heterogeneity of a large family of human LTR-like promoters and enhancers. Nucleic Acids Res *18*, 1261-1270.

Fielder, E., von Zglinicki, T., and Jurk, D. (2017). The DNA Damage Response in Neurons: Die by Apoptosis or Survive in a Senescence-Like State? Journal of Alzheimer's disease : JAD *60*, S107-s131.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell stem cell *2*, 160-169.

Gao, D., Wu, J., Wu, Y.T., Du, F., Aroh, C., Yan, N., Sun, L., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. Science (New York, NY) *341*, 903-906.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Molecular cell *45*, 344-356.

Garcia-Perez, J.L., Widmann, T.J., and Adams, I.R. (2016). The impact of transposable elements on mammalian development. Development (Cambridge, England) *143*, 4101-4114.

Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., and Melnick, A.M. (2015). Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J Vis Exp, e52246.

Garrido-Ramos, M.A. (2017). Satellite DNA: An Evolving Topic. Genes *8*.

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A., and Xu, G.L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem *279*, 25447-25454.

Geiman, T.M., Durum, S.K., and Muegge, K. (1998). Characterization of gene expression, genomic structure, and chromosomal localization of Hells (Lsh). Genomics *54*, 477-483.

Geiman, T.M., Tessarollo, L., Anver, M.R., Kopp, J.B., Ward, J.M., and Muegge, K. (2001). Lsh, a SNF2 family member, is required for normal murine development. Biochim Biophys Acta *1526*, 211-220.

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Research *23*, 789-799.

Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PloS one *5*, e15367.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Golshani, P., Hutnick, L., Schweizer, F., and Fan, G. (2005). Conditional Dnmt1 deletion in dorsal forebrain disrupts development of somatosensory barrel cortex and thalamocortical long-term potentiation. Thalamus Relat Syst *3*, 227-233.

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mobile DNA *7*.

Goto, K., Numata, M., Komura, J.I., Ono, T., Bestor, T.H., and Kondo, H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. Differentiation *56*, 39-44.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem *280*, 13341-13348.

Goyal, R., Reinhardt, R., and Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res *34*, 1182-1188.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nature Reviews Genetics *8*, 35.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011). The role of Tet 3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guo, J.U., Ma, D.K., Mo, H., Ball, M.P., Jang, M.H., Bonaguidi, M.A., Balazer, J.A., Eaves, H.L., Xie, B., Ford, E.*, et al.* (2011). Neuronal activity modifies the DNA methylation landscape in the adult brain. Nature neuroscience *14*, 1345-1351.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G.*, et al.* (2014). Distribution, recognition and regulation of non -CpG methylation in the adult mammalian brain. Nature neuroscience *17*, 215-222.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science (New York, NY) *315*, 1143-1147.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2 -null mutation causes neurological symptoms that mimic Rett syndrome. Nature genetics *27*, 322-326.

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P.*, et al.* (2013). Dynamics of 5 -hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep *3*, 291-300.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877-881.

Han, Y., Ren, J., Lee, E., Xu, X., Yu, W., and Muegge, K. (2017). Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells. Scientific Reports *7*, 1136.

Hancks, D.C., and Kazazian, H.H. (2016). Roles for retrotransp oson insertions in human disease. Mobile DNA *7*.

Hancks, D.C., and Kazazian, H.H., Jr. (2012). Active human retrotransposons: variation and disease. Curr Opin Genet Dev *22*, 191-203.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation -sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486-489.

Hassan, K.M., Norwood, T., Gimelli, G., Gartler, S.M., and Hansen, R.S. (2001). Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication. Human genetics *109*, 452-462.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development (Cambridge, England) *129*, 1983-1993.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S.*, et al.* (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell *6*.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet -mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science (New Yor k, NY) *333*, 1303-1307.

Hendrich, B., and Bickmore, W. (2001). Human diseases with underlying defects in chromatin structure and modification. Hum Mol Genet *10*, 2233-2242.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Molecular and cellular biology *18*, 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes & development *15*, 710-723.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet *27*, 389-396.

Hirai, H., Karian, P., and Kikyo, N. (2011). Regulation of embryonic stem cell self -renewal and pluripotency by leukaemia inhibitory factor. The Biochemical journal *438*, 11-23.

Hoch, R.V., Rubenstein, J.L., and Pleasure, S. (2009). Genes and signaling events that establish regional patterning of the mammalian forebrain. Semin Cell Dev Biol *20*, 378-386.

Hon, G.C., Hawkins, R.D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Hum Mol Genet *18*, R195-201.

Hurst, T.P., and Magiorkinis, G. (2015). Activation of the innate immune response by endogenous retroviruses. Journal of General Virology *96*, 1207-1218.

Hutnick, L.K., Golshani, P., Namihira, M., Xue, Z., Matynia, A., Yang, X.W., Silva, A.J., Schweizer, F.E., and Fan, G. (2009). DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. Hum Mol Genet *18*, 2875-2888.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes & development *29*, 1897-1902.

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences of the United States of America *108*, 3642-3647.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5 -methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, NY) *333*, 1300-1303.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells. Molecular and cellular biology *24*, 8862-8871.

Jair, K.W., Bachman, K.E., Suzuki, H., Ting, A.H., Rhee, I., Yen, R.W., Baylin, S.B., and Schuebel, K.E. (2006). De novo CpG island methylation in human cancer cells. Cancer Res *66*, 682-692.

Jarvis, C.D., Geiman, T., Vila-Storm, M.P., Osipovich, O., Akella, U., Candeias, S., Nathan, I., Durum, S.K., and Muegge, K. (1996). A novel putative helicase produced in early murine lymphocytes. Gene *169*, 203-207.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic -like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum Mol Genet *2*, 731-735.

Jeltsch, A., Broche, J., and Bashtrykov, P. (2018). Molecular Processes Connecting DNA Methylation Patterns with DNA Methyltransferases and Histone Modifications in Mammalian Genomes. Genes *9*.

Jenness, C., Giunta, S., Müller, M.M., Kimura, H., Muir, T.W., and Funabiki, H. ⁀2018⁀. HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proceedings of the National Academy of Sciences.

Jia, J., Shi, Y., Chen, L., Lai, W., Yan, B., Jiang, Y., Xiao, D., Xi, S., Cao, Y., Liu, S.*, et al.* (2017). Decrease in Lymphoid Specific Helicase and 5 -hydroxymethylcytosine Is Associated with Metastasis and Genome Instability. Theranostics *7*, 3920-3932.

Jin, S.G., Wu, X., Li, A.X., and Pfeifer, G.P. (2011). Genomic mapping of 5 -hydroxymethylcytosine in the human brain. Nucleic Acids Res *39*, 5015-5024.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl -cytosine-binding domain links DNA and histone methylation. Current biology : CB *17*, 379-384.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature genetics *19*, 187-191.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes & development *23*, 1303-1312.

Kass, S.U., Goddard, J.P., and Adams, R.L. (1993). Inactive chromatin spreads from a focus of methylation. Molecular and cellular biology *13*, 7372-7379.

Kawasaki, H., Mizuseki, K., Nishikawa, S., Kaneko, S., Kuwana, Y., Nakanishi, S., Nishikawa, S.I., and Sasai, Y. (2000). Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. Neuron *28*, 31-40.

Keshet, I., Yisraeli, J., and Cedar, H. (1985). Effect of regional DNA methylation on gene expression. Proceedings of the National Academy of Sciences of the United States of America *82*, 2560-2564.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M.*, et al.* (2012). 5 -hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nat Struct Mol Biol *19*, 1037-1043.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep *7*, 397-403.

Klein, C.J., Botuyan, M.V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C*., et al.* (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nature genetics *43*, 595-600.

Klug, M., Heinz, S., Gebhard, C., Schwarzfischer, L., Krause, S.W., Andreesen, R., and Rehli, M. (2010). Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels. Genome Biol *11*, R63.

Krajewski, W.A., and Becker, P.B. (1998). Reconstitution of hyperacetylated, DNase I -sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. Proceedings of the National Academy of Sciences of the United States of America *95*, 1540-1545.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5 -hydroxymethylcytosine is present in Purkinje neurons and the brain. Science (New York, NY) *324*, 929-930.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S*., et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS genetics *4*.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994). The DNA (cytosine -5) methyltransferases. Nucleic Acids Res *22*, 1-10.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domai ns that Change during Differentiation. Molecular cell *65*, 432-446.e435.

Kuwabara, T., Hsieh, J., Muotri, A., Yeo, G., Warashina, M., Lie, D.C., Moore, L., Nakashima, K., Asashima, M., and Gage, F.H. (2009). Wnt -mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nature neuroscience *12*, 1097-1105.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R.J., and Defossez, P.A. (2010). The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PloS one *5*, e11982.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis *35*, 88-93.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al.* (2010). Dynamic changes in the human methylome during differentiation. Genome Research *20*, 320-331.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes & development *24*, 265-276.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development (Cambridge, England) *122*, 3195-3205.

Leighton, S.P., Nerurkar, L., Krishnadas, R., Johnman, C., Graham, G.J., and Cavanagh, J. (2017). Chemokines in depression in health and in inflammatory illness: a systematic review and meta-analysis. Molecular Psychiatry *23*, 48.

Lentini, A., Lagerwall, C., Vikingsson, S., Mjoseng, H.K., Douvlataniotis, K., Vogt, H., Green, H., Meehan, R.R., Benson, M., and Nestor, C.E. (2018). A reassessment of DNA -immunoprecipitation-based genomic profiling. Nat Methods *15*, 499-504.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.

Levine, A.J., Singer, E.J., Sinsheimer, J.S., Hinkin, C.H., Papp, J., Dandekar, S., Giovanelli, A., and Shapshak, P. (2009). CCL3 genotype and current depression increase risk of H IV-associated dementia. Neurobehavioral HIV medicine *1*, 1-7.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*.

Li, J.Y., Lees-Murdock, D.J., Xu, G.L., and Walsh, C.P. (2004). Timing of establishment of paternal methylation imprints in the mouse. Genomics *84*, 952-960.

Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature *431*, 364-370.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D.*, et al.* (2013). Global epigenomic reconfiguration during mammalian brain development. Science (New York, NY) *341*, 1237905.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S.*, et al.* (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68-73.

Liu, C., Jiao, C., Wang, K., and Yuan, N. (2018). DNA Methylation and Psychiatric Disorders. Prog Mol Biol Transl Sci *157*, 175-232.

Liu, J., Wu, X., Zhang, H., Pfeifer, G.P., and Lu, Q. (2017). Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. Cell Rep *20*, 1307-1318.

Liu, S., and Tao, Y.G. (2016). Chromatin remodeling factor LSH affects fumarate hydratase as a cancer driver. Chinese Journal of Cancer *35*.

Loo, Y.M., and Gale, M., Jr. (2011). Immune signaling by RIG -I-like receptors. Immunity *34*, 680-692.

Lorenzo, A.D., and Bedford, M.T. (2011). Histone Arginine Methylation. FEBS Lett *585*, 2024-2031.

Lubin, F.D., Roth, T.L., and Sweatt, J.D. (2008). Epigenetic regulation of BDNF gene transcription in the consolidation of fear memory. The Journal of neuroscience : the official journal of the Society for Neuroscience *28*, 10576-10586.

Lucifero, D., Mann, M.R., Bartolomei, M.S., and Trasler, J.M. (2004). Gene -specific timing and epigenetic memory in oocyte imprinting. Hum Mol Genet *13*, 839-849.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L.*, et al.* (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. The EMBO Journal *31*.

Ma, D.K., Jang, M.H., Guo, J.U., Kitabatake, Y., Chang, M., Pow-anpongkul, N., Flavell, R.A., Lu, B., Ming, G., and Song, H. (2009). Neuronal Activity ı Induced Gadd45b Promotes Epigenetic DNA Demethylation and Adult Neurogenesis. Science (New York, NY) *323*, 1074-1077.

Macia, A., Blanco-Jimenez, E., and Garcia-Perez, J.L. (2015). Retrotransposons in pluripotent cells: Impact and new roles in cellular plasticity. Biochim Biophys Acta *1849*, 417-426.

Macia, A., Widmann, T.J., Heras, S.R., Ayllon, V., Sanchez, L., Benkaddour-Boumzaouad, M., Muñoz-Lopez, M., Rubio, A., Amador-Cubero, S., Blanco-Jimenez, E.*, et al.* (2017). Engineered LINE-1 retrotransposition in nondividing human neurons. Genome Research *27*, 335-348.

Mager, D.L., and Stoye, J.P. (2015). Mammalian Endogenous Retroviruses. Microbiol Spectr *3*, Mdna3-0009-2014.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics *2*, e2.

Mallo, M., and Alonso, C.R. (2013). The regulatio n of Hox gene expression during animal development. Development (Cambridge, England) *140*, 3951-3963.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat -associated histone lysine methylation states in the mouse epigenome. The EMBO Journal *24*, 800-812.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Martin Caballero, I., Hansen, J., Leaford, D., Pollard, S., and Hendrich, B.D. (2009). The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. PloS one *4*, e4315.

Martin, G.R. (1981). Isolation of a pluripotent cel l line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation -related chromatin remodeling in activity-dependent BDNF gene regulation. Science (New York, NY) *302*, 890-893.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Research *23*, 1256.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McGhee, J.D., and Felsenfeld, G. (1980). The number of charge -charge interactions stabilizing the ends of nucleosome DNA. Nucleic Acids Res *8*, 2751-2769.

McGraw, C.M., Samaco, R.C., and Zoghbi, H.Y. (2011). Adult neural function requires MeCP2. Science (New York, NY) *333*, 186.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis -regulatory regions. Nat Biotechnol *28*, 495-501.

Medstrand, P., Landry, J.R., and Mager, D.L. (20 01). Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J Biol Chem *276*, 1896-1903.

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res *20*, 5085-5092.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*.

Meehan, R.R., Pennings, S., and Stancheva, I. (2001). Lashings of DNA methylation, forkfuls of chromatin remodeling. Genes & development *15*, 3231-3236.

Meersseman, G., Pennings, S., and Bradbury, E.M. (1992). Mobile nucleosomes --a general behavior. The EMBO Journal *11*, 2951-2959.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high -resolution DNA methylation analysis. Nucleic Acids Res *33*, 5868-5877.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008a). Genome -scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008b). Genome -scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417-1430.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.

Miedel, C.J., Patton, J.M., Miedel, A.N., Miedel, E.S., and Levenson, J.M. (2017). Assessment of Spontaneous Alternation, Novel Object Recognition and Limb Clasping in Transgenic Mouse Models of Amyloid-β and Tau Neuropathology. J Vis Exp.

Migeon, B.R. (2016). An overview of X inactivation based on species differences. Semin Cell Dev Biol *56*, 111-116.

Mignone, J.L., Kukekov, V., Chiang, A.-S., Steindler, D., and Enikolopov, G. (2004). Neural stem and progenitor cells in nestin-GFP transgenic mice. Journal of Comparative Neurology *469*, 311-324.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome -wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage -specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443-446.

Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers an d the transcription cycle. Transcription *2*, 244-253.

Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Molecular and cellular biology *28*, 215-226.

Myant, K., Termanis, A., Sundaram, A.Y., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Research *21*, 83-94.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science (New York, NY) *322*, 1717-1720.

Najjar, S., Pearlman, D.M., Alper, K., Najjar, A., and Devinsky, O. (2013). Neuroinflammation and psychiatric illness. Journal of Neuroinflammation *10*, 816.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M.*, et al.* (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes & development *22*, 37-49.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature *486*, 415-419.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl -CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Nestor, C., Ruzov, A., Meehan, R.R., and Dunican, D.S. (2 010). Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. BioTechniques *48*, 317-319.

Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., and Meehan, R.R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Research *22*, 467-477.

Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E., Schanen, N.C., Zappella, M.*, et al.* (2010). Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol *68*, 944-950.

Ng, H.H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Molecular and cellular biology *20*, 1394-1406.

Ng, R.K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W., and Hemberger, M. (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nature cell biology *10*, 1280-1290.

Nguyen, M.V., Du, F., Felice, C.A., Shan, X., Nigam, A., Mandel, G., Robinson, J.K., and Ballas, N. (2012). MeCP2 is critical for maintaining mature neuronal networks and global brain anatomy during late stages of postnatal brain development and in the mature adult brain. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 10021-10034.

Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., and Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. Dev Dyn *236*, 1663-1676.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Molecular and cellular biology *21*, 4330-4336.

O'Neill, L.A.J., Golenbock, D., and Bowie, A.G. (2013). The history of Toll -like receptors | redefining innate immunity. Nature Reviews Immunology *13*, 453.

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl -CpG binding domain of human MBD1 in complex with methylated DNA. Cell *105*, 487-497.

Okano, H., and Temple, S. (2009). Cell types to order: temporal specification of CNS stem cells. Curr Opin Neurobiol *19*, 112-119.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine -5) methyltransferases. Nature genetics *19*, 219-220.

Oliveira, A.M., Hemstedt, T.J., and Bading, H. (2012). Rescue of aging -associated decline in Dnmt3a2 expression restores cognitive abilities. Nature neuroscience *15*, 1111-1113.

Olivier, I.S., Cacabelos, R., and Naidoo, V. (2018). Risk Factors and Pathogenesis of HIV -Associated Neurocognitive Disorder: The Role of Host Genetics. International journal of molecular sciences *19*.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Oshiumi, H., Miyashita, M., Okamoto, M., Morioka, Y., Okabe, M., Matsumoto, M., and Seya, T. (2015). DDX60 Is Involved in RIG -I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. Cell Reports *11*, 1193-1207.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Current biology : CB *10*, 475-478.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep *10*, 1235-1241.

Pace, J.K., 2nd, and Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Research *17*, 422-432.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes & development *10*, 1991-2002.

Paridaen, J.T., and Huttner, W.B. (2014). Neurogenesis during development of the vertebrate central nervous system. EMBO Rep *15*, 351-364.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO Journal *23*, 4061-4071.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. The Biochemical journal *126*, 781-790.

Peter, I.S., and Davidson, E.H. (2016). Implications of Developmental Gene Regulatory Networks Inside and Outside Developmental Biology. Curr Top Dev Biol *117*, 237-251.

Pham, A.M., Santa Maria, F.G., Lahiri, T., Friedman, E., Marié, I.J., and Levy, D.E. (2016). PKR Transduces MDA5 -Dependent Signals for Type I IFN Induction. PLoS Pathog *12*, e1005489.

Philippidou, P., and Dasen, J.S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. Neuron *80*, 12-34.

Pollard, S.M., Benchoua, A., and Lowell, S. (2006). Neural stem cells, neurons, and glia. Methods Enzymol *418*, 151-169.

Prada, D., González, R., Sánchez, L., Castro, C., Fabián, E., and Herrera, L.A. ⌢2012⌣. Satellite 2 demethylation induced by 5-azacytidine is associated with missegregation of chromosomes 1 and 16 in human somatic cells. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *729*, 100-105.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes & development *15*, 1613-1618.

Prokhortchouk, A., Sansom, O., Selfridge, J., Caballero, I.M., Salozhin, S., Aithozhina, D., Cerchietti, L., Meng, F.G., Augenlicht, L.H., Mariadason, J.M.*, et al.* (2006). Kaiso-deficient mice show resistance to intestinal cancer. Molecular and cellular biology *26*, 199-208.

Qian, X., Shen, Q., Goderie, S.K., He, W., Capela, A., Davis, A.A., and Temple, S. (2000). Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. Neuron *28*, 69-80.

Raabe, E.H., Abdurrahman, L., Behbehani, G., and Arceci, R.J. (2001). An SNF2 factor involved in mammalian development and cellular proliferation. Dev Dyn *221*, 92-105.

Rae, M.M., and Franke, W.W. (1972). The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. Chromosoma *39*, 443-456.

Rajshekar, S., Yao, J., Arnold, P.K., Payne, S.G., Zhang, Y., Bowman, T.V., Schmitz, R.J., Edwards, J.R., and Goll, M. (2018). Pericentromeric hypomethylation elicits an interferon response in an animal model of ICF syndrome. Elife *7*.

Ramesh, V., Bayam, E., Cernilogar, F.M., Bonapace, I.M., Schulze, M., Riemenschneider, M.J., Schotta, G., and Gotz, M. (2016). Loss of Uhrf1 in neural stem cells leads to activation of retroviral elements and delayed neurodegeneration. Genes & development *30*, 2199-2212.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America *97*, 5237-5242.

Ravin, R., Hoeppner, D.J., Munno, D.M., Carmel, L., Sullivan, J., Levitt, D.L., Miller, J.L., Athaide, C., Panchision, D.M., and McKay, R.D. (2008). Potency and fate specification in CNS stem cell populations in vitro. Cell stem cell *3*, 670-680.

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H.*, et al.* (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biology *14*, 1-17.

Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., and Muegge, K. (2015). The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. Nucleic Acids Res *43*, 1444-1455.

Ren, J., Hathaway, N.A., Crabtree, G.R., and Muegge, K. (2018). Tethering of Lsh at the Oct4 locus promotes gene repression associated with epigenetic changes. Epigenetics *13*, 173-181.

Romanish, M.T. (2007). Repeated R ecruitment of LTR Retrotransposons as Promoters by the Anti-Apoptotic Locus NAIP during Mammalian Evolution. *3*.

Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J., Anastassiadis, K., and Stewart, A.F. (2012). Transposon -mediated BAC transgenesis in human ES cells. Nucleic Acids Res *40*, e150.

Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013). INTERFEROME v2.0: an updated database of annotated interferon -regulated genes. Nucleic Acids Research *41*, D1040-D1046.

Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol *225*, 294-303.

Sado, T., Okano, M., Li, E., and Sasaki, H. (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development (Cambridge, England) *131*, 975-982.

Sakai, Y., Suetake, I., Shinozaki, F., Yamashina, S., and Tajima, S. (2004). Co -expression of de novo DNA methyltransferases Dnmt3a2 and Dnmt3L in gonocytes of mouse embryos. Gene Expr Patterns *5*, 231-237.

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., Yamagiwa, A., Niwa, H., Wakayama, T., and Okano, M. (2010). DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. Current biology : CB *20*, 1452-1457.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol *241*, 172-182.

Sasai, N., and Defossez, P.A. (2009). Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. Int J Dev Biol *53*, 323-334.

Sauvageot, C.M., and Stiles, C.D. (2002). Molecu lar mechanisms controlling cortical gliogenesis. Curr Opin Neurobiol *12*, 244-249.

Sawyer, J.R., Swanson, C.M., Wheeler, G., and Cunniff, C. (1995). Chromosome instability in ICF syndrome: formation of micronuclei from multibranched chromosomes 1 demonstrated by fluorescence in situ hybridization. Am J Med Genet *56*, 203-209.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome -wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America *103*, 1412-1417.

Scarsdale, J.N., Webb, H.D., Ginder, G.D., and Williams, D.C., Jr. (2011). Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res *39*, 6741-6752.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nature cell biology *6*, 73-77.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schrader, A., Gross, T., Thalhammer, V., and Langst, G. (2015). Characteriza tion of Dnmt1 Binding and DNA Methylation on Nucleosomes and Nucleosomal Arrays. PloS one *10*, e0140076.

Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34-57.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Screaton, R.A., Kiessling, S., Sansom, O.J., Millar, C.B., Maddison, K., Bird, A., Clarke, A.R., and Frisch, S.M. (2003). Fas -associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. Proceedings of the National Academy of Sciences of the United States of America *100*, 5211-5216.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Sharma, A., Klein, S.S., Barboza, L., Lohdi, N., and Toth, M. (2016). Principles Governing DNA Methylation during Neuronal Lineage and Subtype Specification. The Journal of Neuroscience *36*, 1711-1722.

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Research *21*, 1592-1600.

Shen, Q., Wang, Y., Dimos, J.T., Fasano, C.A., Phoenix, T.N., Lemischka, I.R., Ivanova, N.B., Stifani, S., Morrisey, E.E., and Temple, S. (2006). The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. Nature neuroscience *9*, 743-751.

Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S.*, et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.

Simon, P., Dupuis, R., and Costentin, J. (1994). Thigmotaxis as an index of anxiety in mice. Influence of dopaminergic transmissions. Behavioural Brain Research *61*, 59-64.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non -coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Solovyov, A., Vabret, N., Arora, K.S., Snyder, A., Funt, S.A., Bajorin, D.F., Rosenberg, J.E., Bhardwaj, N., Ting, D.T., and Greenbaum, B.D. (2018). Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. Cell Rep *23*, 512-521.

Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B.*, et al.* (2012). Extensive somatic L1 retrotransposition in colorectal tumors. Genome Research *22*, 2328-2338.

Song, Y., and Brady, S. (2014). Posttranslational Modifications of Tubulin: Pathways to Functional Diversity of Microtubules, Vol 25.

Soshnikova, N., and Duboule, D. (2009). Epigenetic Temporal Co ntrol of Mouse Hox Genes in Vivo. Science (New York, NY) *324*, 1320-1323.

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology *21*, 1973-1985.

Spiers, H., Hannon, E., Schalkwyk, L.C., Bray, N.J., and Mill, J. (2017). 5 -hydroxymethylcytosine is highly dynamic across human fetal brain development. BMC Genomics *18*, 738.

Sproul, D., and Meehan, R.R. (2013). Genomic insights into cancer -associated aberrant CpG island hypermethylation. Brief Funct Genomics *12*, 174-190.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell -intrinsic initiation of autoimmunity. Cell *134*, 587-598.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1 -mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nature cell biology *9*, 1428-1435.

Sun, L.Q., Lee, D.W., Zhang, Q., Xiao, W., Raabe, E.H., Meeker, A., Miao, D., Huso, D.L., and Arceci, R.J. (2004). Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. Genes & development *18*, 1035-1046.

Sun, Z., Wu, Y., Ordog, T., Baheti, S., Nie, J., Duan, X., Hojo, K., Kocher, J.P., Dyck, P.J., and Klein, C.J. (2014). Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. Epigenetics *9*, 1184-1193.

Suzuki, T., Farrar, J.E., Yegnasubramanian, S., Zahed, M., Suzuki, N., and Arceci, R.J. (2008). Stable knockdown of PASG enhances DNA demethylation but does not accelerate cellular senescence in TIG-7 human fibroblasts. Epigenetics *3*, 281-291.

Szabo, P.E., and Pfeifer, G.P. (2012). H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine. Journal of molecular cell biology *4*, 427-429.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I.*, et al.* (2011). 5 -hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nature neuroscience *14*, 1607-1616.

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. The EMBO Journal *27*, 2681-2690.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5 -methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, NY) *324*, 930-935.

Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K.*, et al.* (2007). DAI (DLM -1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. Nature *448*, 501.

Takeshima, H., Suetake, I., Shimahara, H., Ura, K., Tate, S., and Tajima, S. (2006). Distinct DNA methylation activity of Dnmt3a and Dnmt3b towards naked and nucleosomal DNA. Journal of biochemistry *139*, 503-515.

Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T., Huang, J., Stephens, R.*, et al.* (2011). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. Proceedings of the National Academy of Sciences of the United States of America *108*, 5626-5631.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909-920.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausio, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852-10857.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y.*, et al.* (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nature communications *6*, 7870.

Thomson, J.P., and Meehan, R.R. (2017). The appli cation of genome-wide 5-hydroxymethylcytosine studies in cancer research. Epigenomics *9*, 77-91.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes & development *12*, 3693-3702.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell *10*, 1453-1465.

Toubiana, S., Velasco, G., Chityat, A., Kaindl, A.M., Hershtig, N., Tzur-Gilat, A., Francastel, C., and Selig, S. (2018). Subtelomeric methylation distinguishes between subtypes of Immunodeficiency, Centromeric instability and Facial anomalies syndrome. Human Molecular Genetics, ddy265-ddy265.

Tsai, C.L., Shi, Y., and Tainer, J.A. (2014). How substrate specificity is imposed on a histone demethylase | lessons from KDM2A. Genes & development *28*, 1735-1738.

Tuck-Muller, C.M., Narayan, A., Tsien, F., Smeets, D.F., Sawyer, J., Fiala, E.S., Sohn, O.S., and Ehrlich, M. (2000). DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. Cytogenet Cell Genet *89*, 121-128.

Tyagi, M., Imam, N., Verma, K., and Patel, A.K. (2016). Chromatin remodeler s: We are the drivers!! Nucleus *7*, 388-404.

Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Development *133*, 1183-1192.

Uhlhaas, P.J., and Singer, W. (2011). The development of neural synchrony and large - scale cortical networks during adolescence: relevance for the pathophysiology of schizophrenia and neurodevelopmental hypothesis. Schizophrenia bulletin *37*, 514-523.

UniProt Consortium, T. (2018) . UniProt: the universal protein knowledgebase. Nucleic Acids Res *46*, 2699.

Unoki, M., Funabiki, H., Velasco, G., Francastel, C., and Sasaki, H. (2019). CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. The Journal of clinical investigation *129*, 78-92.

Upton K , R., Gerhardt D , J., Jesuadian J , S., Richardson S , R., Sánchez-Luque F , J., Bodea G , O., Ewing A , D., Salvador-Palomeque, C., van der Knaap M , S., Brennan P , M.*, et al.* (2015). Ubiquitous L1 M osaicism in Hippocampal Neurons. Cell *161*, 228-239.

van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S.*, et al.* (2007). L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet *16*, 1587-1592.

van der Vlag, J., and Otte, A.P. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nature genetics *23*, 474-478.

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., and Francastel, C. (2018a). Comparative methylome analysis of

ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Human Molecular Genetics *27*, 2409-2424.

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.F., Chantalat, S., Picard, C., and Francastel, C. (2018b). Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Hum Mol Genet *27*, 2409-2424.

von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., Tanh, M.D., Otto, A., Muegge, K., Heinemann, U.*, et al.* (2012). The SNF2 -like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. The EMBO Journal *31*, 972-985.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA methylation mutants. Science (New York, NY) *260*, 1926-1928.

Waddington, C.H. (1968). Towards a theoretical biology. Nature *218*, 525-527.

Waddington, C.H. (2012). The epigenotype. 1942. Int J Epidemiol *41*, 10-13.

Walsh, C.P., and Bestor, T.H. (1999). Cytosine methylatio n and mammalian development. Genes & development *13*, 26-34.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature genetics *20*, 116-117.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W.*, et al.* (2014). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979-991.

Wang, R., Shi, Y., Chen, L., Jiang, Y., Mao, C., Yan, B., Liu, S., Shan, B., Tao, Y., and Wang, X. (2015). The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH. Scientific Reports *5*.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics *40*, 897-903.

Watanabe, D., Uchiyama, K., and Hanaoka, K. (2006). Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development. Neuroscience *142*, 727-737.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). In itial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DN A methylation in the human genome. Nature genetics *39*.

Wee Yong, V. (2010). Inflammation in neurological disorders: a help or a hindrance? The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry *16*, 408-420.

Weemaes, C.M.R., van Tol, M.J.D., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C.J.A., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E.*, et al.* (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. European Journal Of Human Genetics *21*, 1219.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z.*, et al.* (2006). A global map of p53 transcription -factor binding sites in the human genome. Cell *124*, 207-219.

Whitehouse, I., Flaus, A., Cairns, B.R., White, M.F., Workman, J.L., and Owen-Hughes, T. (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. Nature *400*, 784-787.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Widschwendter, M., Jiang, G., Woods, C., Muller, H.M., Fiegl, H., Goebel, G., Marth, C., Muller-Holzner, E., Zeimet, A.G., Laird, P.W.*, et al.* (2004). DNA hypomethylation and ovarian cancer biology. Cancer Res *64*, 4472-4480.

Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F.*, et al.* (2012). Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet *21*, 2205-2210.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proceedings of the National Academy of Sciences of the United States of America *81*, 2806-2810.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5 -Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature communications *2*, 241.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a -Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. Science (New York, NY) *329*, 444-448.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M.*, et al.* (2016). DNA methylatio n on N(6) -adenine in mammalian embryonic stem cells. Nature *532*, 329-333.

Wu, Z., Huang, K., Yu, J., Le, T., Namihira, M., Liu, Y., Zhang, J., Xue, Z., Cheng, L., and Fan, G. (2012). Dnmt3a regulates both proliferation and differentiation of mouse neural stem cells. J Neurosci Res *90*, 1883-1891.

Xi, S., Geiman, T.M., Briones, V., Guang Tao, Y., Xu, H., and Muegge, K. (2009). Lsh participates in DNA methylation and silencing of stem cell genes. Stem Cells *27*, 2691-2702.

Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T.M., and Muegge, K. (2007). Lsh controls Hox gene silencing during development. Proceedings of the National Academy of Sciences of the United States of America *104*, 14366-14371.

Xiao, C.L., Zhu, S., He, M., Chen, Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.Q., Luo, F.*, et al.* (2018). N(6) -Methyladenine DNA Modification in the Human Genome. Molecular cell *71*, 306-318.e307.

Xiao, Y., Camarillo, C., Ping, Y., Arana, T.B., Zhao, H., Thompson, P.M., Xu, C., Su, B.B., Fan, H., Ordonez, J.*, et al.* (2014). The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. PloS one *9*, e95875.

Yan, Q., Cho, E., Lockett, S., and Muegge, K. (2003). Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. Molecular and cellular biology *23*, 8416-8428.

Yang, X.J., and Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene *26*, 5310-5318.

Yao, B., Christian, K.M., He, C., Jin, P., Ming, G.-l., and Song, H. (2016). Epigenetic mechanisms in neurogenesis. Nature Reviews Neuroscience *17*, 537.

Ye, B., Liu, B., Hao, L., Zhu, X., Yang, L., Wang, S., Xia, P., Du, Y., Meng, S., Huang, G.*, et al.* (2018). Klf4 glutamylation is required for cell reprogramming and early embryonic development in mice. Nature communications *9*, 1261.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F.*, et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (New York, NY) *356*.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003a). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. Cell *115*, 281-292.

Ying, Q.L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol *21*, 183-186.

Yoder, J.A., Soman, N.S., Verdine, G.L., and Bestor, T.H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol *270*, 385-395.

Young, L.E., and Beaujean, N. (2004). DNA methylation in the preimplantation embry o: the differing stories of the mouse and sheep. Animal reproduction science *82-83*, 61-78.

Yu, H., Su, Y., Shin, J., Zhong, C., Guo, J.U., Weng, Y.-L., Gao, F., Geschwind, D.H., Coppola, G., Ming, G.-l*., et al.* (2015). Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. Nature neuroscience *18*, 836-843.

Yu, W., Briones, V., Lister, R., McIntosh, C., Han, Y., Lee, E.Y., Ren, J., Terashima, M., Leighty, R.M., Ecker, J.R*., et al.* (2014a). CG hypomethylation in Lsh -/- mouse embryonic fibroblasts is associated with de novo H3K4me1 formation and altered cellular plasticity. Proceedings of the National Academy of Sciences of the United States of America *111*, 5890-5895.

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M*., et al.* (2014b). Genome -wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Research *24*, 1613-1623.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.

Zegerman, P., Canas, B., Pappin, D., and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem *277*, 11621-11624.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodele r DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193-205.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Current biology : CB *20*, R780-785.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature *466*, 258-262.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W.*, et al.* (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res *38*, 4246-4253.

Zhou, F.C., Resendiz, M., Lo, C.L., and Chen, Y. (2016). Cell -Wide DNA De-Methylation and Re-Methylation of Purkinje Neurons in the Developing Cerebellum. PloS one *11*, e0162063.

Zhou, R., Han, L., Li, G., and Tong, T. (2009). Senescence delay and repression of p16(INK4a) by Lsh via recrui tment of histone deacetylases in human diploid fibroblasts. Nucleic Acids Res *37*, 5183-5196.

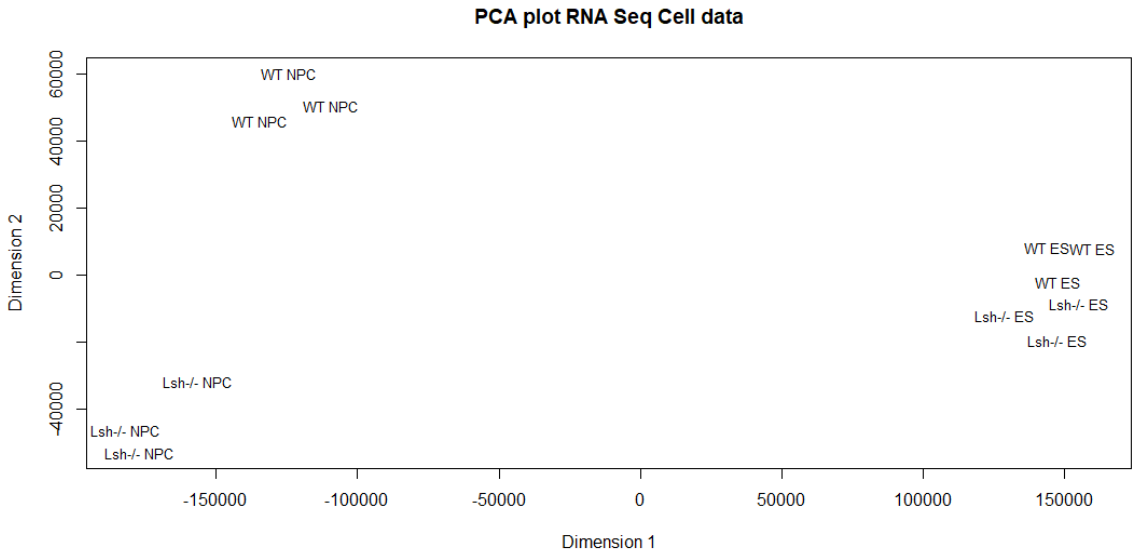
Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription b y inhibiting RNA polymerase II transcriptional elongation. Molecular cell *29*, 69-80.

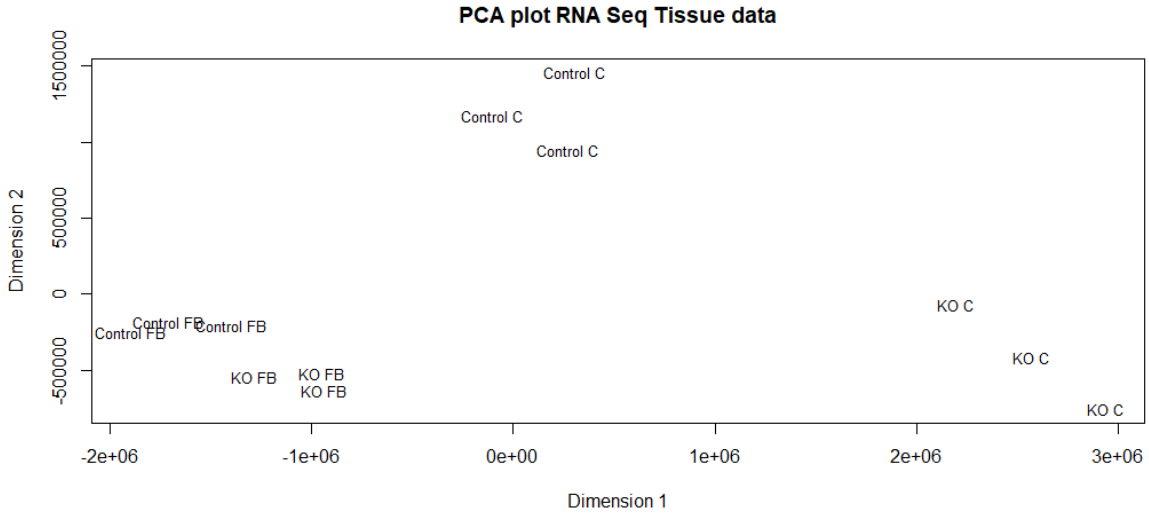
Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. The EMBO Journal *25*, 335-345.

Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T.*, et al.* (2011). Genomic distribution and inter -sample variation of non-CpG methylation across human cell types. PLoS genetics *7*, e1002389.

Appendix

Figure A1





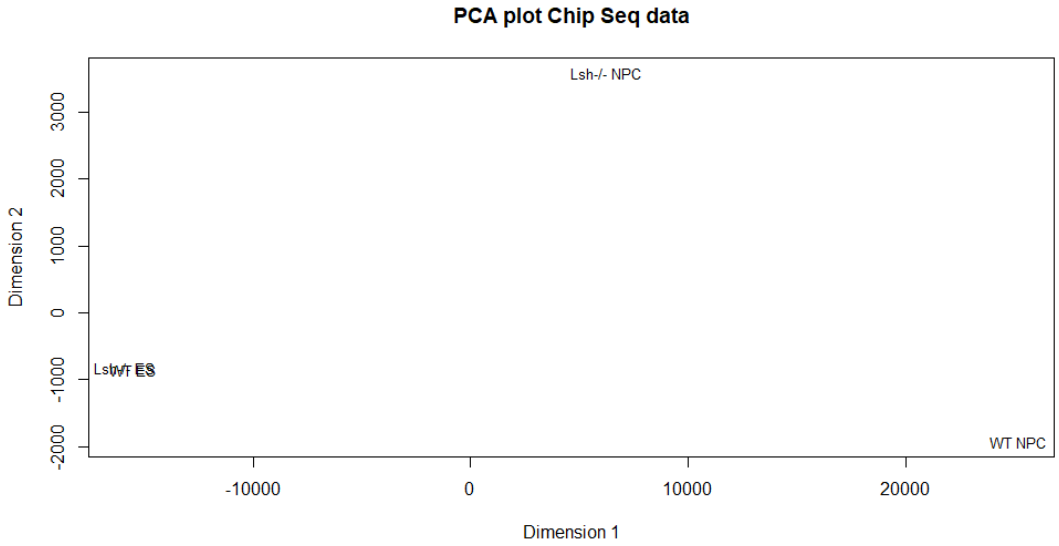


Figure A1. PCA plots from RNA-seq and CHIP-seq data demonstrate that variance is mostly due to cell type ie ES, NPC, forebrain or cerebellum and then due to genotype.

It should be noted that towards the end of this project an off-target mutation was identified in the *Lsh*$^{-/-}$ cells by Dr Ailsa Revuelta. This was a 17bp out of frame deletion in exon 3 of the *Tbkbp1* gene resulting in a premature stop codon. This gene encodes an adaptor protein

which binds to TBK1 and forms part of the interaction network in the TNF/NFκB pathway (Bouwmeester et al., 2004). I therefore looked at expression of this gene in my experimental culture systems by RNA-Seq (Figure A1). This protein is not known to be involved in DNA methylation processes therefore it is unlikely that this would account for the major findings presented in this chapter, however it does highlight the problem of off-target genome editing when using CRISPR technology. In order to confirm that any major findings were indeed due to *Lsh* depletion a rescue experiment was performed with re-expression of *Lsh* in the *Lsh*$^{-/-}$ cell line (section 3.2.4).

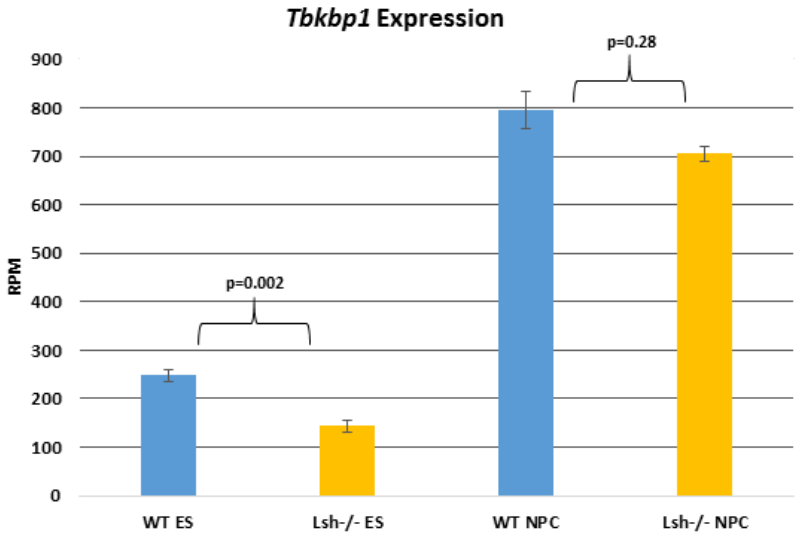


Figure A2. Expression of *Tbkbp1*. Bar chart displays expression of *Tbkbp1* (in which an off target mutation was found) from RNA-Seq analysis on *Lsh*$^{-/-}$ and WT ESCs and NPCs. Data displayed in reads per million aligned reads (RPM). Error bars represent +/- SEM of 3 biological replicates.

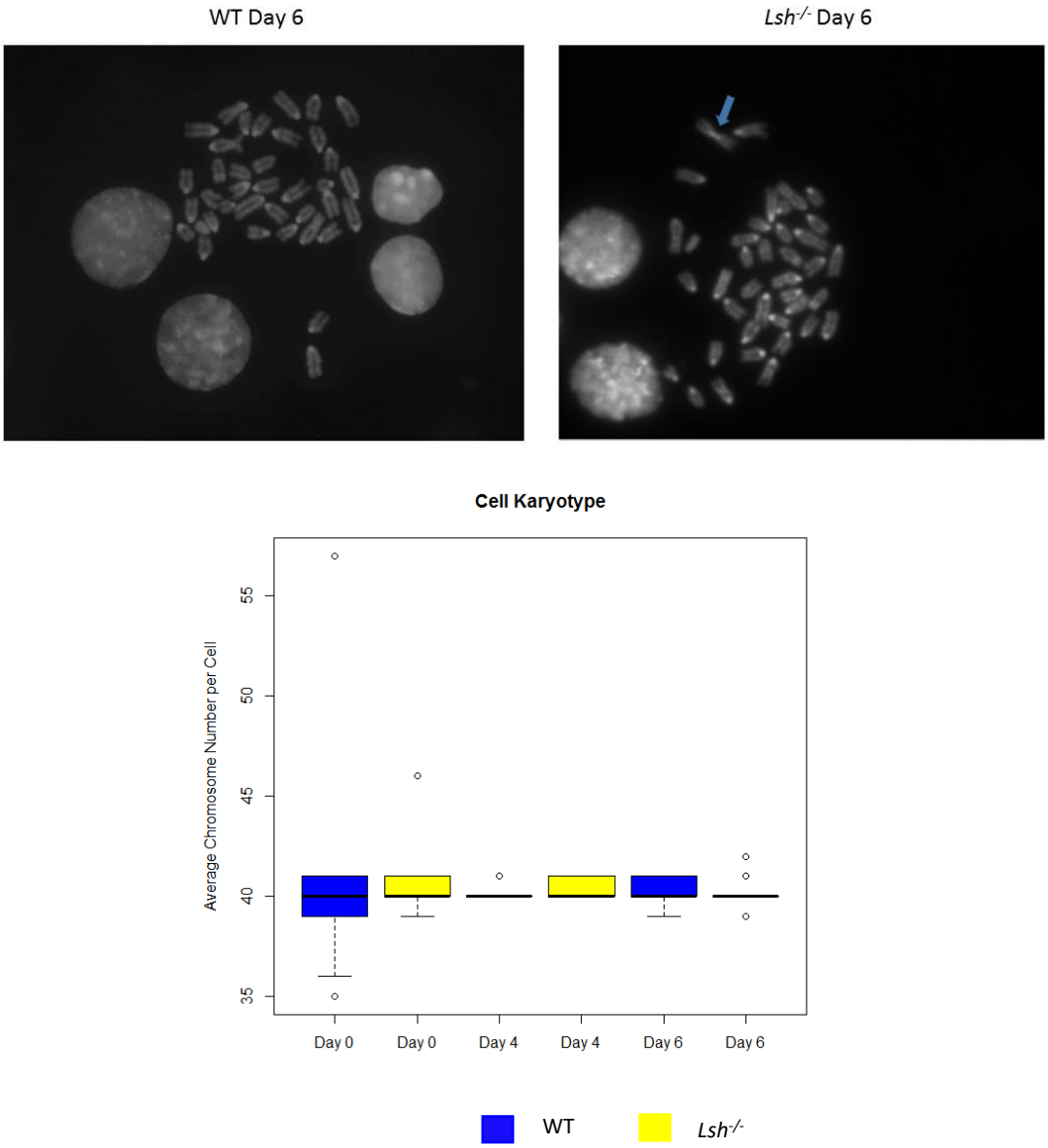
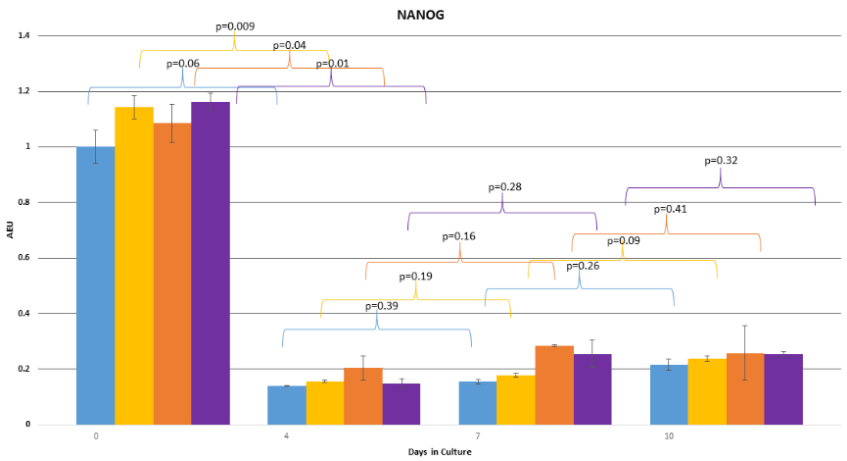
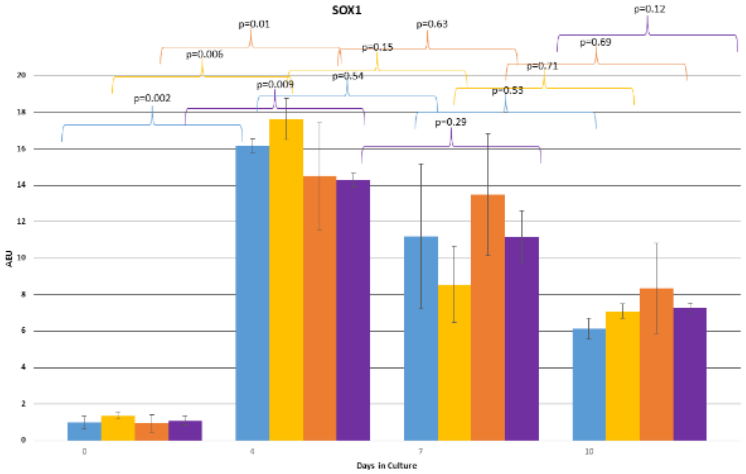
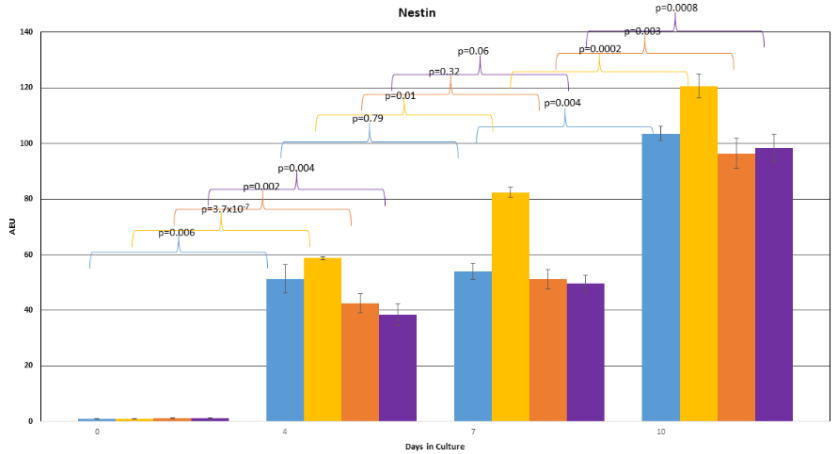
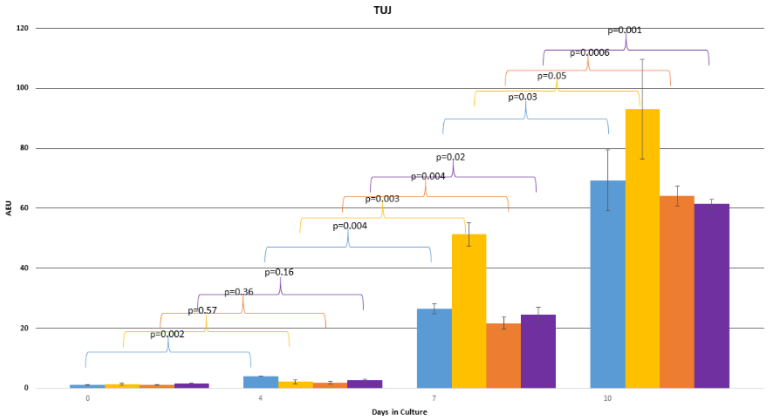


Figure A3. *Lsh*$^{-/-}$ cells show no signs of genomic instability upon neurodifferentiation. Figure shows representative images of metaphase spreads from day 6 of neurodifferentiation protocol. Presence of a robertsonian translocation in *Lsh*$^{-/-}$ cells is indicated by the blue arrow. This translocation was visible from ES cell stage. As can be seen from images there were no other chromosomal abnormalities noted in *Lsh*$^{-/-}$ cells upon differentiation. Boxplot displays the average chromosome number per cell (n=50) during differentiation.


NANOG
p=0.009
p=0.04
p=0.06
p=0.01
p=0.32
p=0.28
p=0.41
p=0.16
p=0.09
p=0.19
p=0.26
p=0.39
AEU
Days in Culture


SOX1
p=0.12
p=0.01
p=0.63
p=0.69
p=0.006
p=0.15
p=0.71
p=0.54
p=0.009
p=0.53
p=0.002
p=0.29
AEU
Days in Culture


Nestin
p=0.0008
p=0.003
p=0.06
p=0.0002
p=0.32
p=0.01
p=0.004
p=0.79
p=0.004
p=0.002
p=3.7x10-7
p=0.006
AEU
Days in Culture


TUJ
p=0.001
p=0.0006
p=0.05
p=0.03
p=0.02
p=0.004
p=0.003
p=0.004
p=0.16
p=0.36
p=0.57
p=0.002
AEU
Days in Culture

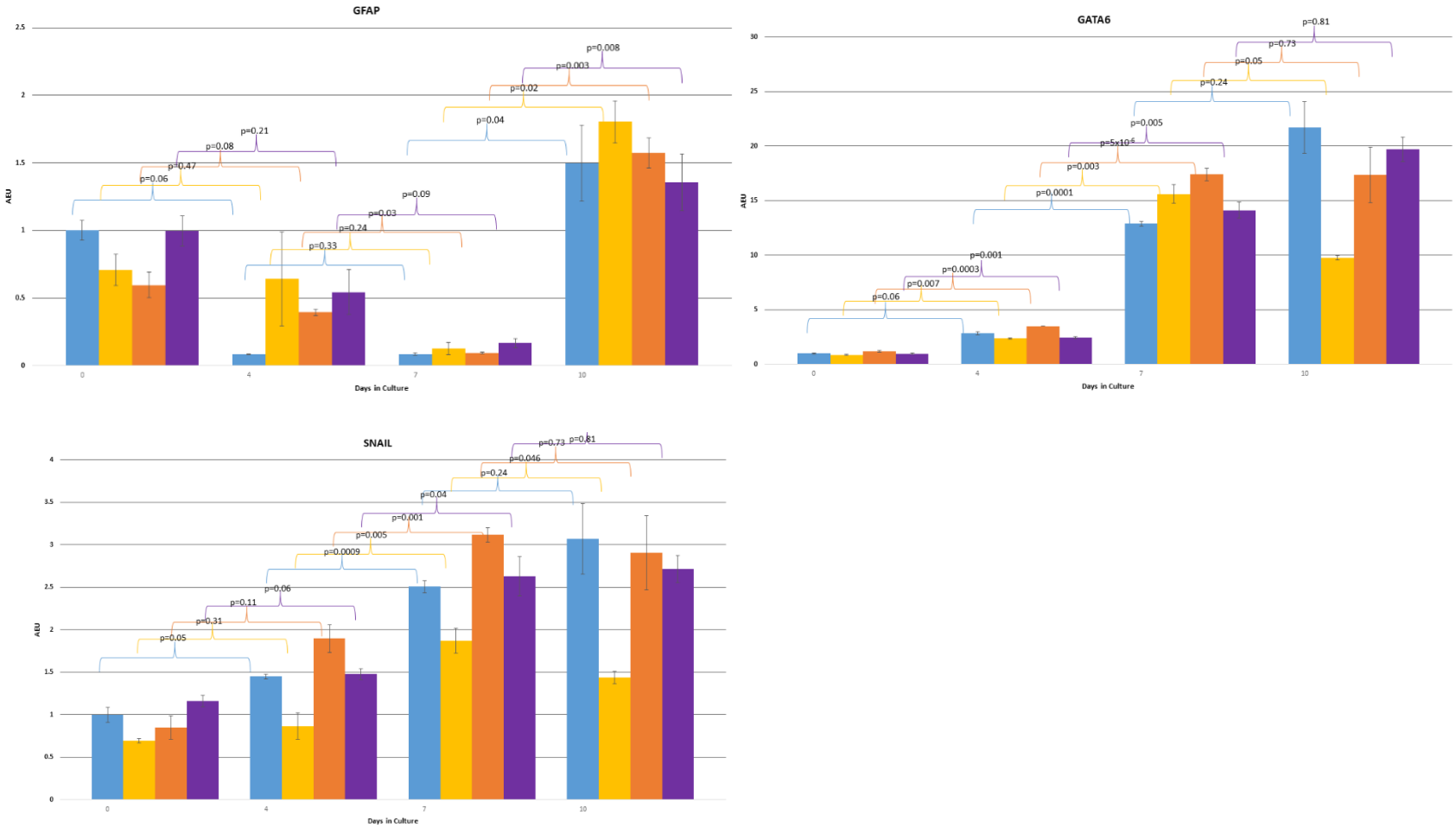
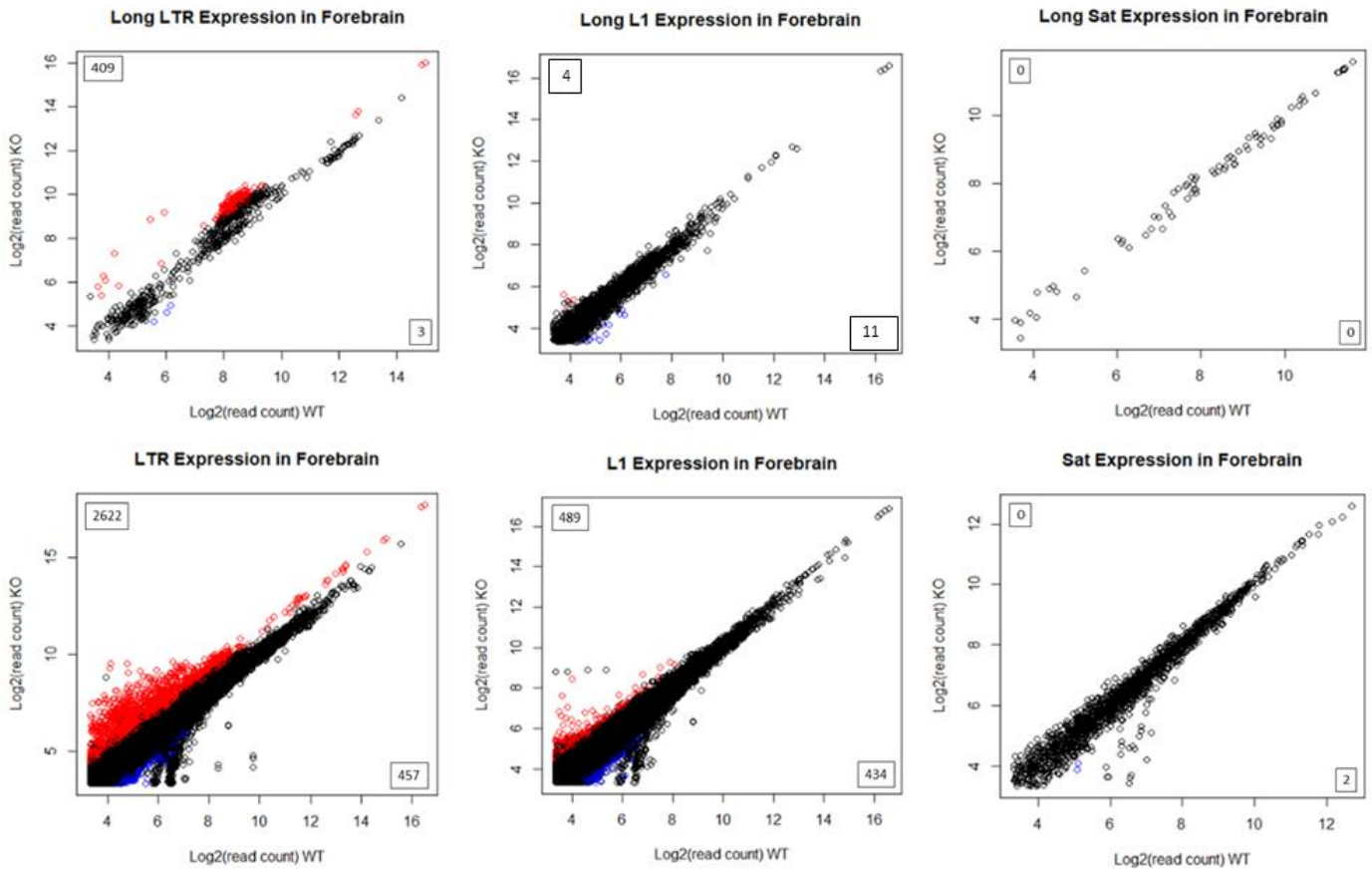


Figure A4. Expression of neural lineage markers during N2B27 differentiation. Quantification of expression of lineage marker genes by qRT-PCR. Expression is normalised to housekeeping *Gapdh* expression and then standardized to WT ES expression. Error bars represent +/- SEM of 3 biological replicates. P-values indicate level of significance calculated by two-tailed student t-test. Only significant values are displayed. AEU=arbitrary expression units.

Figure A5.


Long LTR Expression in Forebrain
409
3
Long L1 Expression in Forebrain
4
11
Long Sat Expression in Forebrain
0
0
LTR Expression in Forebrain
2622
457
L1 Expression in Forebrain
489
434
Sat Expression in Forebrain
0
2
Log2(read count) KO
Log2(read count) WT

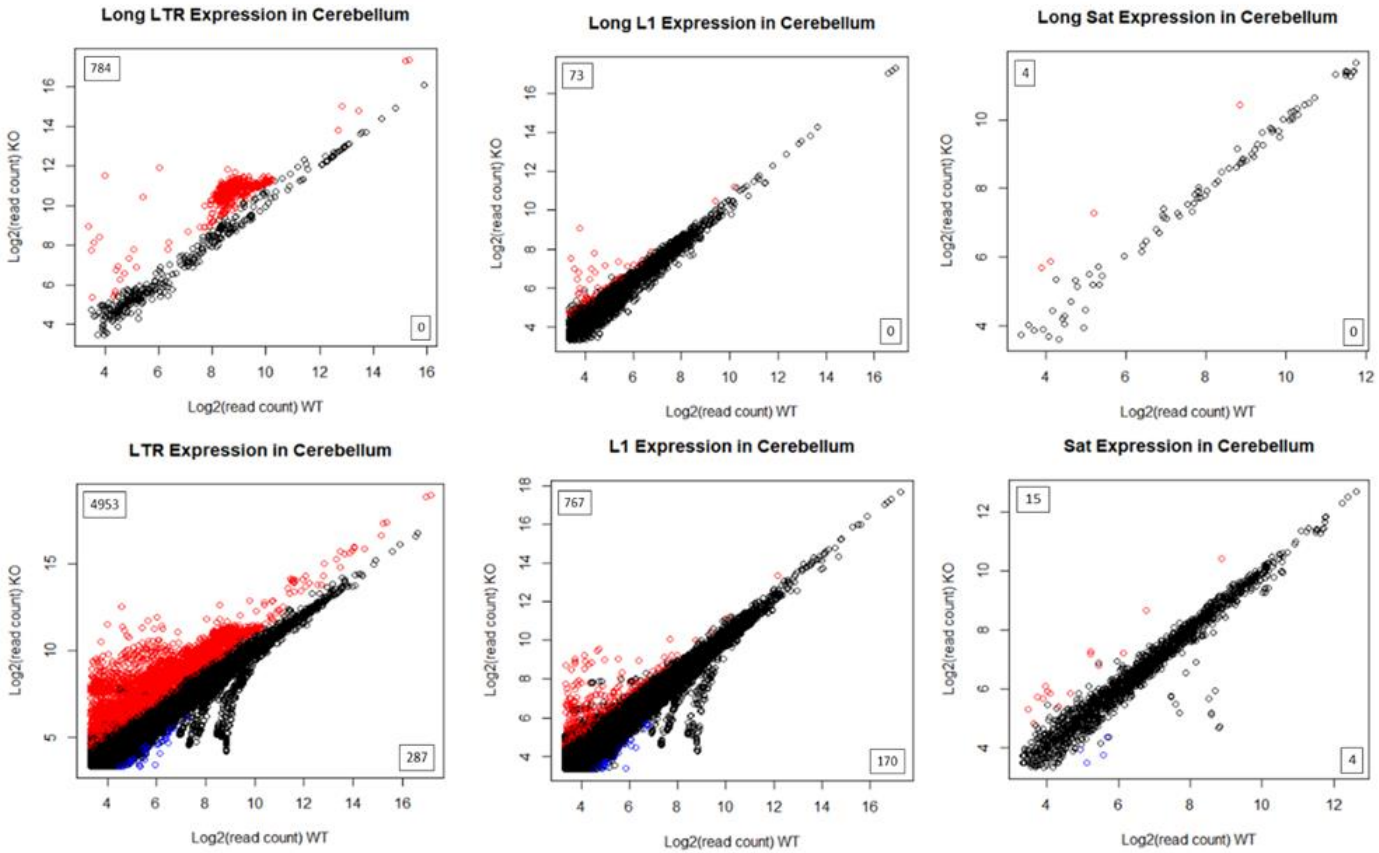


Figure A5 Repeat expression in neural tissue from non-uniquely mapped reads. Scatter plots display read counts of LINE1 (L1), long terminal repeat (LTR) and satellite (Sat) sequences from RNA-Seq analysis from 3 biological replicates of control and KO forebrain tissue. Repeats defined as differentially expressed in KO tissue (>2 fold change and FDR<0.05) are highlighted in red if up-regulated and blue if down-regulated. Data is shown for reads mapping to full length repeats (long) and non-full length repeats (short). This data is generated from reads prior to filtering for uniquely mapping reads only.

***List A1**: Overlapping down-regulated genes in Lsh$^{-/-}$ ES and NP cells*

2200002D01Rik, AC154842.1, Ankrd33b, Atp6ap2, AU015836, AY036118, Cbr3, Ddx3y, Dsc2, Frem2, Gm13227, Hhip, Igf2, Moxd1, Myh9, Nexn, Nup62cl, Pdzk1, Perp, Plac8, Prrg4, Rhox6, Rnf128, Sfmbt2, Sfrp2, Slc23a1, Slc30a3, Slc40a1, Stard8, Synj2, Tek, Uty, Wnt7b

***List A2:** Overlapping up-regulated genes in Lsh$^{-/-}$ ES and NP cells*

AI854703, Bcas1, Grin2c, Miip, Nefm, Nfib, Pla2g4c, Ptgs1, Rftn1, SNORA42.3, Ttll4

***List A3:** Overlapping hypomethylated gene promoters in KO cerebellum, forebrain and NPCs*

4930571K23Rik, 9030619P08Rik, Adam6a, Bsph1, Catsper1, Csta, Cyp2d34, Cyp8b1, Defb7, Gm11213, Gm16501, Gm6026, Gm904, LOC100009614, Mageb16, Nlrp5, Olfr1340, Olfr1415, Olfr1474, Olfr406-ps, Olfr775, P2ry4, Serpina1b, Serpina3c, Serpinb9c, Sult2a4, Tbx19, Usp9y, Vmn1r65

***List A4:** Overlapping hypomethylated gene promoters in KO cerebellum and forebrain*

1700013B16Rik, 3830403N18Rik, Adh4, Aif1, Ccdc108, Chrna2, Cyp1a2, Defb50, Egam-1c, Fcgr4, Gm12888, Gm16378, Iigp1, Klra21, Klra8, Krt6b, Lce3c, Mat1a, Mir1936, Mir673, Morc2b, Nr0b2, Olfr1362, Olfr52, Olfr742, Pcdhb18, Pklr, Pramef12, Prss40, Sfi1, Slc38a4, Speer2, Sycn, Tas2r109, Tectb, Tnfsf14, Vmn2r106, Xlr3c, Xlr4a, Zfp345

***List A5:** Overlapping hypomethylated gene promoters in KO forebrain and NPCs*

Ccl11, Chrm2, Clec4a1, Elf3, Enpep, Heatr7b2, Krt32, Ly6i, Magea8, Olfr796, Ppp1r3a, Tgm4, Vmn2r116

***List A6:** Overlapping hypomethylated gene promoters in KO cerebellum and NPCs*

1700023I07Rik, 1700027A23Rik, 1700108M19Rik, 1700111N16Rik, 1810009J06Rik, 4921523A10Rik, 4930403N07Rik, 4930474M22Rik, 4930562C15Rik, 4933421I07Rik, 5830405N20Rik, A1cf, A630001G21Rik, A930017M01Rik, AF067063, Actl9, Art2a-ps, Art2b, Asb17, BC117090, C130026I21Rik, Ccdc153, Ccdc27, Ccl12, Ccl3, Cd177, Cd200r2, Ceacam2, Ces2f, Cidec, Cldn13, Clec4b2, Csf2rb, Cyp2c38, Cyp4f40, Dbh, Defb29, Defb6,

Dnahc8, Doxl2, Fbxw22, Fcgbp, G6pd2, Gm1006, Gm1027, Gm13251, Gm16387, Gm5065, Gm5072, Gm5077, Gm5416, Gm5736, Gm6904, Gm8096, Gm9731, Gml, Gprc2a-rs5, Gsdma3, Gsdmcl-ps, H2-DMb1, H2-M10.3, H2-M10.4, Hemt1, Hhipl2, Hmgb1-rs17, Hnf4a, Htr3b, Itih1, Klk6, Klkb1, Klra5, Krt1, Krt74, Krt77, Krt84, Krtap10-4, Lcn14, Lcn6, Lpcat2b, Lrrc39, Ly6c2, Magea2, Mir684-2, Mrgprb8, Ms4a5, Ms4a6c, Ms4a7, Mslnl, Mup21, Nkg7, Nlrc5, Nlrp2, Nlrp4c, Oas2, Olfr1342, Olfr1395, Olfr1396, Olfr1416, Olfr1494, Olfr215, Olfr222, Olfr368, Olfr506, Olfr527, Olfr54, Olfr633, Olfr655, Olfr656, Olfr678, Olfr734, Olfr872, Parp10, Pcdha8, Pde6g, Pilra, Pilrb1, Pira6, Psg22, Ptk6, Rbm31y, Rhox2h, Rhox8, Rnf222, Ropn1, Rprl2, Sbpl, Serpina1a, Serpina9, Serpinb3b, Slc34a1, Slpi, Smok2b, Spata3, Spt1, Svs2, Sytl3, Tmem150b, Tmem195, Tmem92-ps, Trcg1, Trim60, Trim61, Trim75, Trpv3, Tsks, Upp2, Vmn1r89, Vmn2r16, Vmn2r69, Vmn2r92, Was, Wfdc10, Xlr4b

List A7: *Interferon regulated genes up-regulated in KO forebrain*

1700030C10Rik, Ahsg, Alb, Ambp, Apoa1, Apoc3, Apol9b, AV099323, BC094916, Bcl3, C920025E04Rik, Casp4, Cbx3, Ccl24, Ccl3, Ccl4, Ccl5, Ccrn4l, Cd2, Cd52, Cdkn2a, Ces1c, Clec7a, Cst7, Cxcl10, Cyp2a5, Cyp2f2, Cyp3a11, Ddx60, F2, Fabp1, Fbp1, Fev, Fga, Fgb, Fgg, Gbp10, Gbp6, Gc, Gm10068, Gm16340, Gm2619, Gm4841, Gm4951, Gm5431, Gm7609, H2-Q10, H2-Q5, Hal, Hamp2, Hpx, Ifi204, Ifi27l2a, Ifitm1, Itgax, Kng1, Lgals3, Mug1, Mup10, Mup11, Nr5a1, Pcp2, Pet2, Pigr, Pira2, Pyhin1, Pzp, Rgn, Sectm1a, Serpina3k, Serpinb1-ps1, Serpinb1c, Serpinc1, Slc44a3, Slc6a4, Trim5, Tspan32, U6, Xist, Zbp1

List A8: *Interferon regulated genes up-regulated in KO cerebellum*

2010005H15Rik, A530032D15Rik, AC168977.1, Actn3, Ahsg, Akap14, Alb, Apoc3, Apoh, Bst2, C130026I21Rik, Ccnb1, Ccrn4l, Cd209f, Cd79b, Cdcp1, Cdk6, Cks1b, Cryba4, Cxcl10, Ddx60, Dpep1, Epcam, Fam167b, Fgb, Gbp10, Gbp6, Gc, Gdf15, Glipr1, Gm11127, Gm11427, Gm12856, Gm15753, Gm6264, Gm6531, Gp49a, Hamp2, Hp, Ifi205, Ifi27l2a, Ifi44, Irf7, Itgax, Lilra6, Mndal, Mx1, Ncapg, Nr5a1, Oas2, Pet2, Pira2, Plbd1, Scn5a, Sectm1a, Serpinb1-ps1, Serpinb1c, Soat2, Sp140, Spc25, Trim5, U6, Ugt3a2, Usp18, Zbp1

List A9: *Overlapping genes up-regulated in KO cerebellum and forebrain*

A430078G23Rik, AC110247.1, AC151987.1, AC155816.1, AC155933.1, Ahsg, AL672068.1, Alb, Apoc3, Btnl7, Ccrn4l, CT033780.1, Cxcl10, Ddx60, E030037K03Rik, E430029J22Rik, Eps8l1, Fetub, Fgb, Gbp10, Gc, Ghrh, Gm10040, Gm11382, Gm14149, Gm15930, Gm5751,

Gm6969, Gm9864, Gstp2, Hamp2, Heatr7b1, Ifi27l2a, Igsf1, Itgax, Klra5, Ly9, Mrgprx2, Mup17, Mup20, Mup3, Mup7, Nr5a1, Nxf3, Olfr1054, P2rx3, Pcdhb6, Pet2, Pira2, Sectm1a, Serpina1c, Serpinb1c, Strc, Trim5, Usp9y, Zbp1

List A10: *Up-regulated genes in KO cerebellum with hypomethylated promoters*

A430078G23Rik, Aurkc, Bglap-rs1, Brs3, C130026I21Rik, Cd79b, Dpep1, Fgb, Gstp2, Klra5, Krt28, Ly9, Oas2, Pcdhb13, Pcdhb2, Pira2, Serpina1c, Usp9y, Xlr3c, Xlr4a, Xlr4b, Zbp1

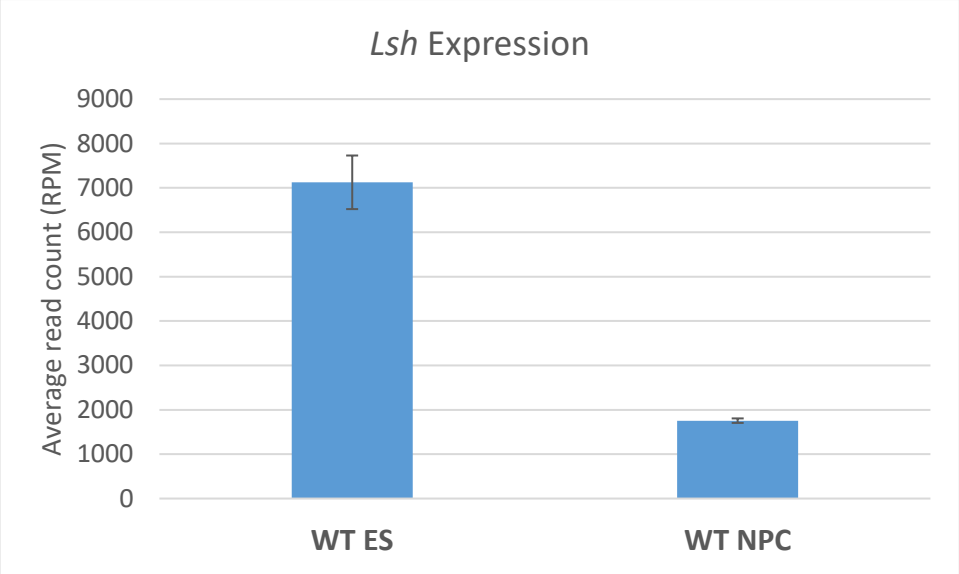


Figure A6. *Lsh* is down-regulated upon differentiation. Bar chart displays expression of *Lsh* from RNA-Seq analysis on WT ES and NP cells. Data displayed in reads per million aligned reads (RPM). Error bars represent +/- SEM of 3 biological replicates.

Gene	**Repeat class (distance from TSS bp)**
Up-regulated in KO Cerebellum	
Ccrn41	LTR short(157,300) LTR short(20,645)
Cpxm2	LTR short(77,322) LTR short (82,822) LTR short (83,069)
Eps811	Sat short(11,377)
Gm10800	Sat short(363)
Gm11127	LTR short(2428) LTR long(5798) LTR short(9168)
Gm14406	LTR short(26,451) L1 short(25,741) L1 short (9364) LTR short (11,246) LTR short (11,956) LTR short (12,658)
Gm14444	L1 short(10,421)
Hamp2	LTR short(3413)
Lilrb3	LTR Long(47,208)
Mrgprx2	LTR short(19,523)
Olfr922	L1 short(3037)
Pcdhb10	LTR short (4423)
Pcdhb11	LTR short (5321)
Serpinb1c	LTR short(5823) L1 short (22,625) LTR short(23,296)
Strc	LTR short(2150)
Wdr31	LTR short(23,656)
Up-regulated in KO forebrain	
Cyp2a5	LTR long(72,894)

Table A1. Up-regulated genes in KO tissue with de-repressed repeats in regulatory domains. Table displays up-regulated genes in KO mouse neural tissue from RNA-Seq analysis that contain an up-regulated repeat within their regulatory domain as defined by default settings of the genomic regions enrichment of annotations tool (GREAT) (McLean et al., 2010). The class of the up-regulated repeat(s) in the domain and whether they are full length (long) or not (short) are displayed with the distance from the transcription start site, in bp, displayed in parenthesis.

Table A2

P53 target genes	Forebrain				Cerebellum			
	log2(FC)	P value	control average RPM	KO average RPM	log2(FC)	P value	control average RPM	KO average RPM
adora2b	0.401969	0.012989	140.3333	191	0.130501	0.585321	147	161.3333
adrb1	-0.21859	0.048353	1391	1210.667	0.021232	0.918609	199	202.3333
ank1	-0.1017	0.29711	4288	4062.667	-0.1182	0.301726	28049	26210.67
ankrd10	-0.12018	0.232002	2404.667	2284.333	0.082947	0.405917	2658.667	2846
arhgap5	0.122623	0.266386	4948.667	5580.667	-0.28451	0.066028	8217.667	6912.667
astn2	-0.16777	0.114765	2143	1960.667	0.071235	0.674341	6045.667	6508
atf3	0.26455	0.247804	65	78.66667	-0.15158	0.555403	86	79
bax	-0.02469	0.803855	1433.333	1438.667	0.151871	0.115872	1843.333	2062
bcas3	-0.05158	0.583658	2438.667	2409.667	-0.02141	0.82468	2374	2358.667
bcl2a1	1.241078	0.07809	2.666667	6.666667	0.570144	0.457859	4	6
bicd2	-0.11925	0.201284	4562	4306	-0.07072	0.406241	5241	5045
mmadhc	0.023862	0.809622	1448.333	1498.667	-0.1026	0.284063	1897	1784
cald1	0.140236	0.179911	2191.667	2492.667	0.034832	0.721756	3348.667	3447.667
cblc	0.334759	0.79562	1	1.333333	0.930237	0.366899	1.666667	3.333333
ccng2	-0.05439	0.581594	1844.333	1819.333	-0.18535	0.249156	3153	2829
cdc42ep3	0.078257	0.472404	620.3333	672.3333	-0.03147	0.8325	533.6667	527
cdkal1	0.037549	0.747007	459	488	-0.09405	0.421292	635.3333	601.3333
cdkn1a	0.177988	0.108937	734	845.3333	0.297357	0.031026	775.3333	957.3333
chd2	-0.10253	0.269994	3451.667	3318.333	-0.08016	0.405691	5464.333	5235.667
chst12	-0.16314	0.185355	754.3333	689	0.062579	0.690094	697.3333	731
ncapd2	-0.24213	0.01751	1396	1210.333	-0.01311	0.922168	2704.333	2706.667
col4a1	0.046812	0.61513	4154	4381	0.375135	0.006479	2398.667	3128.667
ctnna3	-0.22469	0.239824	81.66667	72.33333	-0.50244	0.016826	322	232.3333
ddb2	-0.17125	0.115209	793.6667	728.3333	0.054254	0.726412	1212	1265.667

ddit4	-0.02647	0.803698	1909	1921.667	-0.22292	0.061533	2586.333	2220.667
dkk3	-0.19475	0.05502	12448	11136.67	0.07547	0.713762	1728.667	1818.333
eea1	-0.04629	0.632311	2746.667	2751.333	-0.19702	0.037246	3484.333	3079.333
eif2ak	-0.02963	0.742698	3281.333	3312.333	-0.12296	0.172183	3707.667	3436.667
erbb4	-0.00016	0.998884	3188.333	3295	-0.26556	0.071509	1885.667	1596.667
fat	-0.15271	0.163741	3080.333	2831.333	-0.21166	0.225247	13077.33	11588.33
fgf2	-0.21505	0.338645	50	44.33333	-0.26395	0.349053	49	41.66667
frmd4a	0.076014	0.407849	2474.333	2690	0.035292	0.703784	2727.333	2820.333
gadd45a	0.224221	0.09932	284.3333	338.3333	0.157704	0.240415	429.3333	481
gml	0	1	0	0	1.866178	0.344034	0	0.333333
gnai1	0.02218	0.812579	6975	7344	-0.32022	0.003813	7683.333	6237.333
gnaq	0.08363	0.370581	9508.333	10476	-0.10524	0.24914	9186	8615.333
gpr39	0.23589	0.676625	6	7.333333	0.75153	0.370653	2.333333	4
gspt1	0.084473	0.376742	5193	5708.333	-0.20893	0.031832	6028	5279.667
hdac9	-0.02066	0.830819	1719	1757.333	-0.20955	0.179305	771.3333	671.6667
ier5	-0.13736	0.177834	3206.333	2950	0.039264	0.689202	1148.667	1187
itgam	0.300321	0.009839	1081.333	1367	0.09618	0.603127	434.3333	471
kcnma1	-0.33294	0.005569	2968	2444.333	-0.26933	0.30234	2155.667	1856.667
lats2	-0.10357	0.273233	1760.333	1687	-0.15457	0.167513	3073.667	2809
ltbp1	0.182837	0.127	519.3333	602.6667	-0.01728	0.916313	318	316
mdm4	-0.22472	0.043106	2025	1792	-0.18404	0.220136	2793	2516.667
mlh1	-0.09854	0.57639	441.3333	442.6667	0.083328	0.749995	606	642.6667
msh6	-0.10119	0.350581	1088.333	1033.667	-0.08253	0.434161	1646.333	1577.333
mybl1	-0.03739	0.797559	213.6667	214.3333	-0.30083	0.055832	522	430
myo1a	-0.36596	0.072393	74.33333	59.33333	-0.06597	0.774931	129.3333	125
nab1	-0.0336	0.726308	1788.333	1803.333	-0.05262	0.598773	2884.333	2811.667
nav3	-0.15537	0.108365	3217	2987	-0.10688	0.441884	960.3333	907.6667
nck2	-0.04236	0.66536	1244.333	1244.333	0.027739	0.803196	961	992.3333
nedd4l	-0.03184	0.72825	6492	6571	-0.01301	0.92137	5153.667	5201.333

neo1	-0.15116	0.149136	5045	4718.333	-0.10098	0.335774	4815.667	4553
nid2	0.325388	0.013737	306.6667	397.3333	0.535868	0.005387	235.6667	345.6667
nlgn1	-0.02904	0.796764	2784	2841	0.04029	0.765447	1257.333	1315.667
nmu	-1.7035	0.255935	1.333333	0.333333	-2.68404	0.025323	11.33333	1.666667
notch1	-0.10815	0.254189	2066	1983	-0.02223	0.807172	1955	1943.333
nr6a1	-0.25653	0.038133	569.3333	497.3333	-0.17678	0.41067	549.3333	494
osbp	-0.10684	0.243318	3834	3670.333	-0.08372	0.315399	4024	3837.667
pcca	0.023797	0.807153	1425	1503	-0.05026	0.686185	2953.667	2883.333
pcdh7	-0.44825	0.000454	2919	2181	-0.4645	0.10005	1150.667	841.3333
pcna	-0.01413	0.885255	1645.333	1694.333	-0.15242	0.087579	1940	1759
phf14	-0.02778	0.767777	2203.667	2227	-0.04893	0.56221	2077.667	2029
pias2	-0.02098	0.825667	2729.333	2790.333	-0.20375	0.033034	3335.667	2933
ppm2c	-0.03269	0.741919	6030.667	6103	-0.19327	0.068492	2327.333	2060.333
prdm1	0.088098	0.660348	72.66667	78.33333	0.070634	0.796596	40.33333	42.66667
prkag2	-0.00892	0.922082	3675.333	3784	0.02426	0.827628	3757.333	3858.667
pstpip2	-0.50909	0.001176	274	200	-0.49733	0.052337	283.6667	201
ptk2	-0.15629	0.090962	6055	5594.667	-0.07257	0.37588	3466.667	3325.667
ptpre	-0.12635	0.196561	1737.333	1649	0.013181	0.908948	2538.333	2602
ptprm	-0.09468	0.33182	2453.333	2350.333	-0.02257	0.860753	3422.333	3429.333
ptpro	-0.10127	0.477283	1973	1846.667	0.48426	0.122865	496	715.6667
rbl1	0.234735	0.196169	127.6667	156	0.018313	0.921212	313.3333	322
rps27l	0.367893	0.00519	376.6667	502.3333	0.231794	0.15478	489.6667	574.6667
rrad	0.184621	0.302363	84.66667	99.33333	0.197061	0.484389	83.33333	97.66667
rrm2b	0.090235	0.387858	1600	1768	-0.08864	0.425166	1420.333	1344.667
shrml	-0.271	0.048124	332.6667	281.6667	-0.18574	0.21472	577.6667	514
slc4a10	-0.14147	0.220329	11021.33	10325.33	-0.10089	0.508174	8311.667	7937.333
smarcb1	-0.05496	0.569695	2918.667	2884	0.022572	0.80221	3308.333	3383.667
snk	-0.04492	0.617641	7237.667	7240.333	0.015571	0.879359	1319.667	1344
snx5	0.160064	0.099527	3944	4553	-0.0311	0.726544	4166.667	4103.667

spag9	-0.03216	0.741107	19450.33	19620.67	-0.1931	0.106298	21310	18925.33
stard4	-0.09663	0.344787	1629	1562.667	-0.31533	0.004684	2739.667	2232
stau	0.044566	0.622176	3333.333	3543	-0.07199	0.404225	3782.667	3623.333
tgfa	-0.18848	0.049111	2138.333	1942	-0.37578	0.026939	1491.667	1151.667
tif1	-0.00524	0.955597	1721.333	1762	-0.07941	0.367912	1835.333	1748.667
tnfaip8	0.404967	0.00225	323.6667	439.3333	0.086879	0.539507	378	405
tnfrsf10b	-0.04797	0.78944	94.33333	93.66667	0.223758	0.279885	99.66667	118
tpo	0.170101	0.590352	30.66667	34.33333	-0.28265	0.341132	29.66667	24.66667
trpm1	-0.02111	0.945334	23.33333	24	-0.04811	0.875773	21.66667	21.33333
ubp1	-0.02845	0.761555	4349.333	4426.333	0.005493	0.963948	6434.667	6519.667
usp34	-0.08805	0.402447	8145.333	7981.667	-0.21834	0.062732	9121.667	7981
usp9x	0	1	0	0	0	1	0	0
vim	0.002794	0.975489	2736	2815.333	0.087499	0.445684	7380.333	7945
wig1	-0.03721	0.679981	4071.333	4085.667	-0.09063	0.345319	4183.667	3962.333
xpc	-0.09163	0.353693	1151	1107.333	0.124427	0.190963	1239.667	1360

Table A2. Expression of p53 target genes in mouse neural tissue. Table displays RNA-Seq data of expression of p53 target genes from control and KO mouse and cerebellum. Average RPM is the average read count per million over 3 replicates. Highlighted are those genes considered differentially expressed i.e., displaying >2fold change in expression and p value <0.05.